GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 14:31:01 ; Search time 40 Seconds
(without alignments)
38.467 Million cell updates/sec

Title: US-09-763-848-1
Perfect score: 99
Sequence: 1 PEWPSYLGYEKLGPYY 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	99	100.0	284	2	S68216	phosphoprotein pho
2	52	52.5	406	2	S42394	G-box-binding prot
3	52	52.5	700	2	S38361	calpain (EC 3.4.22
4	51	51.5	349	2	A41349	histone-specific t
5	51	51.5	349	2	S77570	transcription fact
6	51	51.5	700	1	CIHUH2	calpain (EC 3.4.22
7	49	49.5	705	1	CICEH	calpain (EC 3.4.22
8	49	49.5	842	2	T23715	hypothetical prote
9	49	49.5	1160	2	T23713	hypothetical prote
10	49	49.5	1286	2	T23714	hypothetical prote
11	48	48.5	127	2	B31807	Ig heavy chain V r
12	48	48.5	601	2	H90159	hypothetical prote
13	47.5	48.0	273	2	D71221	probable morphine
14	47	47.5	245	2	AG0701	Orf 245 protein [i
15	47	47.5	301	1	A48041	protein kinase (EC
16	47	47.5	456	2	T49997	hypothetical prote
17	47	47.5	566	2	T03855	hypothetical prote
18	47	47.5	703	2	A48764	calpain (EC 3.4.22
19	46	46.5	327	2	D83320	hypothetical prote
20	46	46.5	340	1	C64096	aldose 1-epimerase
21	45	45.5	344	2	T00110	integral membrane
22	45	45.5	385	2	T09880	omega-6 desaturase
23	45	45.5	611	1	FPRT	alpha-fetoprotein
24	45	45.5	700	2	S57194	calpain (EC 3.4.22
25	44.5	44.9	177	2	B72580	hypothetical prote
26	44.5	44.9	421	2	AC1481	conserved hypothet
27	44.5	44.9	421	2	AH1120	B. subtilis YwbN p
28	44	44.4	268	2	S08229	chlorophyll a/b-bi
29	44	44.4	401	2	T52250	probable alanine-g
30	44	44.4	554	2	AE0584	asparagine synthet
31	44	44.4	761	2	A53414	A-kinase anchor pr
32	43.5	43.9	359	2	F82316	probable extracell
33	43	43.4	120	2	S22958	adenosylhomocystei
34	43	43.4	267	2	T15645	hypothetical prote
35	43	43.4	378	1	QXBY33	oxi3 intron 3 prot
36	43	43.4	415	2	S78646	DNA endonuclease I
37	43	43.4	421	2	S33998	K421R protein - Af
38	43	43.4	503	2	C86250	hypothetical prote
39	43	43.4	544	2	T27444	hypothetical prote
40	43	43.4	730	2	D87365	prolyl oligopeptid
41	43	43.4	1094	2	T05472	hypothetical prote
42	43	43.4	1296	2	I40645	botulinum neurotox
43	43	43.4	2970	2	T08839	polyprotein - marm
44	42.5	42.9	401	2	A82221	extracellular solu
45	42.5	42.9	735	2	S58551	photosystem I prot

ALIGNMENTS

RESULT 1
S68216
phosphoprotein phosphatase (EC 3.1.3.16) 1 glycogen-binding chain - rat
N;Alternate names: 33K protein glycogen-binding chain (G(L))
C;Species: Rattus norvegicus (Norway rat)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C;Accession: S68216; S74276; S68723
R;Doherty, M.J.; Moorhead, G.; Morrice, N.; Cohen, P.; Cohen, P.T.W.
FEBS Lett. 375, 294-298, 1995
A;Title: Amino acid sequence and expression of the hepatic glycogen-binding (G(L))-s
A;Reference number: S68216; MUID:96085228; PMID:7498521
A;Accession: S68216
A;Molecule type: mRNA
A;Residues: 1-284 <DOH>
A;Cross-references: GB:S80360; NID:g1245930; PID:g1245931
A;Accession: S74276
A;Molecule type: protein
A;Residues: 21-24;43-56;61-98;150-157;161-169;181-199;201-220;238-253;280-284 <DOE>
R;Moorhead, G.; MacKintosh, C.; Morrice, N.; Cohen, P.
FEBS Lett. 362, 101-105, 1995
A;Title: Purification of the hepatic glycogen-associated form of protein phosphatase
A;Reference number: S68721; MUID:95237359; PMID:7720853
A;Accession: S68723
A;Molecule type: protein
A;Residues: 61-76;150-157 <MOO>
A;Experimental source: liver
C;Keywords: glycogen metabolism; phosphoric monoester hydrolase
F;1-284/Product: phosphoprotein phosphatase 1 glycogen-binding chain #status experi

Query Match 100.0%; Score 99; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 PEWPSYLGYEKLGPYY 16
              ||||||||||||||||
Db        269 PEWPSYLGYEKLGPYY 284
```

RESULT 2
S42394
G-box-binding protein - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C;Accession: S42394
R;Meier, I.; Gruissem, W.
Nucleic Acids Res. 22, 470-478, 1994
A;Title: Novel conserved sequence motifs in plant G-box binding proteins and implic
A;Reference number: S42392; MUID:94173701; PMID:8127687
A;Accession: S42394
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-406 <MEI>

A;Cross-references: EMBL:X74943; NID:g456754; PIDN:CAA52897.1; PID:g456755
C;Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology
F;286-326/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 52.5%; Score 52; DB 2; Length 406;
Best Local Similarity 46.7%; Pred. No. 2.6;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPY 15
|||| : || : |:
Db 25 PEWPGFQGYSPMPPH 39

RESULT 3
S38361
calpain (EC 3.4.22.17) II large chain - rat
N;Alternate names: calpain II 80K chain
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 22-Jun-1999
C;Accession: S38361; S08650; S39751
R;DeLuca, C.I.; Davies, P.L.; Samis, J.A.; Elce, J.S.
Biochim. Biophys. Acta 1216, 81-93, 1993
A;Title: Molecular cloning and bacterial expression of cDNA for rat calpain II 80 kDa su
A;Reference number: S38361; MUID:94032492; PMID:8218419
A;Accession: S38361
A;Molecule type: mRNA
A;Residues: 1-700 <DEL>
A;Cross-references: EMBL:L09120; NID:g402665; PIDN:AAA16327.1; PID:g402666
R;Samis, J.A.; Back, D.W.; Graham, E.J.; Elice, J.S.
submitted to the EMBL Data Library, February 1990
A;Reference number: S08650
A;Accession: S08650
A;Molecule type: DNA
A;Residues: 380-439 <SAM>
A;Cross-references: EMBL:X51772
C;Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C;Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer; hyd
F;75-327/Domain: calpain catalytic domain homology <CALP>
F;529-560/Domain: calmodulin repeat homology <EF1>
F;572-604/Domain: calmodulin repeat homology <EF2>
F;605-634/Domain: calmodulin repeat homology <EF3>
F;637-669/Domain: calmodulin repeat homology <EF4>
F;105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 52.5%; Score 52; DB 2; Length 700;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPY 15
| || ||:::|||
Db 52 PALPSSLGFKELGPY 66

RESULT 4
A41349
histone-specific transcription factor HBP1 - wheat
C;Species: Triticum *sp.* (wheat)
C;Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 09-Jun-2000
C;Accession: A41349
R;Tabata, T.; Takase, H.; Takayama, S.; Mikami, K.; Nakatsuka, A.; Kawata, T.; Nakayama,
Science 245, 965-967, 1989
A;Title: A protein that binds to a cis-acting element of wheat histone genes has a leuci
A;Reference number: A41349; MUID:89368924; PMID:2772648
A;Accession: A41349
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-349 <TAB>
A;Cross-references: GB:M28704; NID:g170748; PIDN:AAA34293.1; PID:g170749
C;Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription regulation
F;247-287/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 51.5%; Score 51; DB 2; Length 349;
Best Local Similarity 46.7%; Pred. No. 3.2;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPY 15
|||| : || : |:
Db 34 PEWPGFQGYPAMPPH 48

RESULT 5
S77570
transcription factor HBP-1a(17) - wheat
N;Alternate names: basic leucine zipper protein HBP-1a(17)
C;Species: Triticum aestivum (common wheat)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S77570; S58693; S15346
R;Mikami, K.
submitted to the EMBL Data Library, August 1994
A;Reference number: S77570
A;Accession: S77570
A;Molecule type: DNA
A;Residues: 1-349 <MIK>
A;Cross-references: EMBL:D38111; NID:g1199789; PIDN:BAA07289.1; PID:g1199790
R;Mikami, K.; Katsura, M.; Ito, T.; Okada, K.; Shimura, Y.; Iwabuchi, M.
Mol. Gen. Genet. 248, 573-582, 1995
A;Title: Developmental and tissue-specific regulation of the gene for the wheat basi
A;Reference number: S58693; MUID:96027929; PMID:7476857
A;Accession: S58693
A;Molecule type: mRNA
A;Residues: 1-18 <MIW>
R;Tabata, T.; Nakayama, T.; Mikami, K.; Iwabuchi, M.
EMBO J. 10, 1459-1467, 1991
A;Title: HBP-1a and HBP-1b: leucine zipper-type transcription factors of wheat.
A;Reference number: S15346; MUID:91224097; PMID:2026143
A;Accession: S15346
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-349 <TAB>
A;Cross-references: EMBL:X56781; NID:g21632; PIDN:CAA40101.1; PID:g21633
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 199
C;Genetics:
A;Introns: 18/3; 40/3; 68/3; 96/3; 114/3; 128/1; 147/2; 164/3; 177/1; 248/3; 274/3;
C;Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology
C;Keywords: DNA binding; leucine zipper; nucleus; transcription factor
F;247-287/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 51.5%; Score 51; DB 2; Length 349;
Best Local Similarity 46.7%; Pred. No. 3.2;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPY 15
|||| : || : |:
Db 34 PEWPGFQGYPAMPPH 48

RESULT 6
CIHUH2
calpain (EC 3.4.22.17) large chain 2 - human
N;Alternate names: calpain chain L-2; calpain II catalytic chain; high-calcium requi
C;Species: Homo sapiens (man)
C;Date: 21-Nov-1993 #sequence_revision 09-Aug-1997 #text_change 16-Jul-1999
C;Accession: S10590; A31218; A33529
R;Sorimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saido, T.C.; Ohno, S.; Minami, Y.
Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990
A;Title: A novel member of the calcium-dependent cysteine protease family.
A;Reference number: S10589; MUID:90380278; PMID:2400579
A;Accession: S10590
A;Molecule type: mRNA
A;Residues: 1-700 <SOR>
R;Imajoh, S.; Aoki, K.; Ohno, S.; Emori, Y.; Kawasaki, H.; Sugihara, H.; Suzuki, K.
Biochemistry 27, 8122-8128, 1988
A;Title: Molecular cloning of the cDNA for the large subunit of the high-Ca(2+)-requi
A;Reference number: A31218; MUID:89166474; PMID:2852952
A;Accession: A31218

A;Molecule type: mRNA; protein
A;Residues: 1-210,'I',212-394,'D',396-445,'I',447-700 <IMA>
A;Cross-references: GB:M23254; NID:g511636; PIDN:AAA35645.1; PID:g511637
A;Note: parts of this sequence were determined by protein sequencing; the amino end of t
R;Hata, A.; Ohno, S.; Akita, Y.; Suzuki, K.
J. Biol. Chem. 264, 6404-6411, 1989
A;Title: Tandemly reiterated negative enhancer-like elements regulate transcription of a
A;Reference number: A33529; MUID:89197947; PMID:2539381
A;Accession: A33529
A;Molecule type: DNA
A;Residues: 1-67,'G',69-72,'IE',75-78,'R' <HAT>
A;Cross-references: DDBJ:J04700; NID:g179910; PIDN:AAA52760.1; PID:g463086
C;Genetics:
A;Gene: GDB:CAPN2; mCANP; CANPml
A;Cross-references: GDB:119750; OMIM:114230
A;Map position: 1pter-1qter
C;Complex: heterodimer of L (large) and S (small) chains
C;Function:
A;Description: catalyzes the hydolysis of peptides
A;Note: cleaves preferentially after tyrosine, methionine, or arginine residues and befo
C;Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C;Keywords: acetylated amino end; calcium binding; cysteine proteinase; EF hand; heteroc
F;2-700/Product: calpain large chain 2 #status predicted <MAT>
F;75-327/Domain: calpain catalytic domain homology <CALP>
F;529-560/Domain: calmodulin repeat homology <EF1>
F;572-604/Domain: calmodulin repeat homology <EF2>
F;605-634/Domain: calmodulin repeat homology <EF3>
F;637-669/Domain: calmodulin repeat homology <EF4>
F;2/Modified site: acetylated amino end (Ala) (in mature form) #status predicted
F;105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 51.5%; Score 51; DB 1; Length 700;
Best Local Similarity 60.0%; Pred. No. 6.7;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66
```

RESULT 7
CICHE
calpain (EC 3.4.22.17) large chain 4 - chicken
N;Alternate names: calpain catalytic chain; intermediate calcium activated neutral prote
C;Species: Gallus gallus (chicken)
C;Date: 17-May-1985 #sequence_revision 09-Aug-1997 #text_change 24-Nov-1999
C;Accession: A00979
R;Ohno, S.; Emori, Y.; Imajoh, S.; Kawasaki, H.; Kisaragi, M.; Suzuki, K.
Nature 312, 566-570, 1984
A;Title: Evolutionary origin of a calcium-dependent protease by fusion of genes for a th
A;Reference number: A93348; MUID:85061606; PMID:6095110
A;Accession: A00979
A;Molecule type: mRNA
A;Residues: 1-705 <OHN>
A;Cross-references: EMBL:X01415; NID:g63332; PIDN:CAA25658.1; PID:g63333
R;Emori, Y.; Ohno, S.; Tobita, M.; Suzuki, K.
FEBS Lett. 194, 249-252, 1986
A;Reference number: A91354; MUID:86082358; PMID:3000828
A;Contents: annotation; gene structure
C;Comment: This calpain has calcium requirements intermediate between those of the high-
C;Genetics:
A;Introns: 82/3; 106/1; 145/3; 190/2; 246/3; 274/3; 303/2; 328/2; 382/1; 438/3; 442/3; 5
C;Complex: heterodimer of L (large) and S (small) chains
C;Function:
A;Description: catalyzes the hydolysis of peptides
A;Note: cleaves preferentially after tyrosine, methionine, or arginine residues and befo
C;Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C;Keywords: blocked amino end; calcium binding; cysteine proteinase; duplication; EF han
F;78-330/Domain: calpain catalytic domain homology <CALP>
F;533-564/Domain: calmodulin repeat homology <EF1>
F;576-608/Domain: calmodulin repeat homology <EF2>
F;609-638/Domain: calmodulin repeat homology <EF3>
F;641-673/Domain: calmodulin repeat homology <EF4>
F;674-705/Domain: calmodulin repeat homology <EF5>
F;2/Modified site: blocked amino end (Met) (in mature form) #status experimental
F;108,265,289/Active site: Cys, His, Asn #status predicted

Query Match 49.5%; Score 49; DB 1; Length 705;
Best Local Similarity 53.3%; Pred. No. 14;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 PEWPSYLGYEKLGPY 15
              |  |: ||:::|||| 
Db         55 PAGPTALGFKELGPY 69
```

RESULT 8
T23715
hypothetical protein M04C9.6c - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23715
R;Burton, J.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19787
A;Accession: T23715
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-842 <WIL>
A;Cross-references: EMBL:Z83731; PIDN:CAB54273.1; GSPDB:GN00019; CESP:M04C9.6c
A;Experimental source: clone M04C9
C;Genetics:
A;Gene: CESP:M04C9.6c
A;Map position: 1
A;Introns: 71/3; 91/3; 114/3; 262/2; 322/2; 374/1; 500/1; 587/2; 645/2; 833/2

Query Match 49.5%; Score 49; DB 2; Length 842;
Best Local Similarity 42.9%; Pred. No. 16;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 EWPSYLGYEKLGPY 15
              :| |:: |:::|||: 
Db        145 KWVSFIAYQQLGPF 158
```

RESULT 9
T23713
hypothetical protein M04C9.6b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23713
R;Burton, J.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19787
A;Accession: T23713
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1160 <WIL>
A;Cross-references: EMBL:Z83731; PIDN:CAB06024.1; GSPDB:GN00019; CESP:M04C9.6b
A;Experimental source: clone M04C9
C;Genetics:
A;Gene: CESP:M04C9.6b
A;Map position: 1
A;Introns: 71/3; 91/3; 114/3; 262/2; 322/2; 374/1; 461/2; 519/2; 707/2; 727/3; 768/

Query Match 49.5%; Score 49; DB 2; Length 1160;
Best Local Similarity 42.9%; Pred. No. 23;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 EWPSYLGYEKLGPY 15
              :| |:: |:::|||: 
Db        145 KWVSFIAYQQLGPF 158
```

RESULT 10
T23714

```
hypothetical protein M04C9.6a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23714
R;Burton, J.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19787
A;Accession: T23714
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1286 <WIL>
A;Cross-references: EMBL:Z83731; PIDN:CAB06025.1; GSPDB:GN00019; CESP:M04C9.6a
A;Experimental source: clone M04C9
C;Genetics:
A;Gene: CESP:M04C9.6a
A;Map position: 1
A;Introns: 71/3; 91/3; 114/3; 262/2; 322/2; 374/1; 500/1; 587/2; 645/2; 833/2; 853/3; 89

  Query Match            49.5%;  Score 49;  DB 2;  Length 1286;
  Best Local Similarity  42.9%;  Pred. No. 26;
  Matches    6;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 EWPSYLGYEKLGPY 15
            :| |:: |::|||:
Db      145 KWVSFIAYQQLGPF 158


RESULT 11
B31807
Ig heavy chain V region (PAC1) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Jul-1989 #sequence_revision 20-Jul-1989 #text_change 23-May-1997
C;Accession: B31807
R;Taub, R.; Gould, R.J.; Garsky, V.M.; Ciccarone, T.M.; Hoxie, J.; Friedman, P.A.; Shatt
J. Biol. Chem. 264, 259-265, 1989
A;Title: A monoclonal antibody against the platelet fibrinogen receptor contains a seque
A;Reference number: A31807; MUID:89079661; PMID:2909518
A;Accession: B31807
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-127 <TAU>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;15-97/Domain: immunoglobulin homology <IMM>

  Query Match            48.5%;  Score 48;  DB 2;  Length 127;
  Best Local Similarity  61.5%;  Pred. No. 3.2;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 PSYLGYEKLGPYY 16
              |||  |:  ||||
Db        100 PSYYRYDGAGPYY 112


RESULT 12
H90159
hypothetical protein lig [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C;Accession: H90159
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90159
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-601 <KUR>
A;Cross-references: GB:AE006641; NID:g13813323; PIDN:AAK40535.1; GSPDB:GN00155
C;Genetics:
A;Gene: lig
C;Superfamily: DNA ligase

  Query Match            48.5%;  Score 48;  DB 2;  Length 601;
  Best Local Similarity  63.6%;  Pred. No. 16;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 WPSYLGYEKLG 13
              || :||| :||
Db         50 WPDFLGYPELG 60


RESULT 13
D71221
probable morphine 6-dehydrogenase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: D71221
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; S
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: D71221
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <KAW>
A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29099.1; PID:g3256416
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenB
C;Genetics:
A;Gene: PH0031
C;Superfamily: aldehyde reductase

  Query Match            48.0%;  Score 47.5;  DB 2;  Length 273;
  Best Local Similarity  41.7%;  Pred. No. 8.6;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels   11;  Gaps    1;

Qy          3 WPSYLGYEK----------LGPY 15
              |||: ||||          || |
Db         85 WPSHPGYEKAKKAAKASAKRLGTY 108


RESULT 14
AG0701
Orf 245 protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (stra
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AG0701
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churc
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AG0701
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-245 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD01984.1; PID:g16502824; GSPDB:GN00176
C;Genetics:
A;Gene: STY1741

  Query Match            47.5%;  Score 47;  DB 2;  Length 245;
  Best Local Similarity  72.7%;  Pred. No. 9.1;
  Matches    8;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 EWPSYLGYEKL 12
              |||  ||| ||
Db        172 EWPPLLGYNKL 182
```

```
RESULT 15
A48041
protein kinase (EC 2.7.1.37) cdc2-related CRK1 - Leishmania mexicana
C;Species: Leishmania mexicana
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A48041; S31366
R;Mottram, J.C.; Kinnaird, J.H.; Shiels, B.R.; Tait, A.; Barry, J.D.
J. Biol. Chem. 268, 21044-21052, 1993
A;Title: A novel CDC2-related protein kinase from Leishmania mexicana, LmmCRK1, is post-
A;Reference number: A48041; MUID:94012652; PMID:8407941
A;Accession: A48041
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-301 <MOT>
A;Cross-references: EMBL:X60385; NID:g9539; PIDN:CAA42936.1; PID:g9540
C;Superfamily: kinase-related transforming protein; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;3-257/Domain: protein kinase homology <KIN>
F;11-19/Region: protein kinase ATP-binding motif
F;34,52,127,129/Active site: Lys, Glu, Asp, Lys #status predicted

  Query Match             47.5%;  Score 47;  DB 1;  Length 301;
  Best Local Similarity   58.8%;  Pred. No. 11;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    4;  Gaps    1;

Qy          1 PEWPSYL----GYEKLG 13
              ||| : |    ||||||
Db        251 PEWSNVLGSVPGYEKLG 267


Search completed: October  2, 2003, 14:36:38
Job time : 41 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 14:23:45 ; Search time 22 Seconds
(without alignments)
34.201 Million cell updates/sec

Title: US-09-763-848-1
Perfect score: 99
Sequence: 1 PEWPSYLGYEKLGPYY 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	56	56.6	700	1	CAN2_MOUSE	Q08529	mus musculu
2	52	52.5	700	1	CAN2_RAT	Q07009	rattus norv
3	51	51.5	349	1	HBPA_WHEAT	P23922	triticum ae
4	51	51.5	700	1	CAN2_HUMAN	P17655	homo sapien
5	49	49.5	705	1	CANX_CHICK	P00789	gallus gall
6	48	48.5	600	1	DNLI_SULTO	Q976g4	sulfolobus
7	48	48.5	601	1	DNLI_SULSH	Q9p9k9	sulfolobus
8	48	48.5	601	1	DNLI_SULSO	Q980t8	sulfolobus
9	47	47.5	301	1	CRK1_LEIME	Q06309	leishmania
10	46	46.5	340	1	GALM_HAEIN	P31765	haemophilus
11	45	45.5	611	1	FETA_RAT	P02773	rattus norv
12	45	45.5	700	1	CAN2_CHICK	Q92178	gallus gall
13	45	45.5	748	1	SUF2_HUMAN	Q9hau4	homo sapien
14	44	44.4	268	1	CB2_MALDO	P15773	malus domes
15	44	44.4	415	1	LEU2_PYRAE	Q8zw41	pyrobaculum
16	44	44.4	687	1	AKA8_RAT	Q63014	rattus norv
17	43	43.4	120	1	SAHH_STRFR	P26799	streptomyce
18	43	43.4	378	1	AI3M_YEAST	P03877	saccharomyc
19	43	43.4	421	1	V421_ASFB7	Q07384	african swi
20	43	43.4	482	1	SAHH_STRAA	Q936d6	streptomyce
21	43	43.4	1295	1	BXA2_CLOBO	Q45894	clostridium
22	42.5	42.9	735	1	PSAB_MAIZE	P04967	zea mays (m
23	42	42.4	174	1	CHCB_BOMMO	P08830	bombyx mori
24	42	42.4	196	1	SODF_TETPY	P19666	tetrahymena
25	42	42.4	619	1	SUF1_MOUSE	Q9cun6	mus musculu
26	42	42.4	702	1	CANB_HUMAN	Q9umq6	homo sapien
27	42	42.4	714	1	CAN1_HUMAN	P07384	homo sapien
28	42	42.4	731	1	SUF1_XENLA	Q9pun2	xenopus lae
29	42	42.4	757	1	SUF1_HUMAN	Q9hce7	homo sapien
30	42	42.4	1295	1	BXA1_CLOBO	P10845	clostridium
31	41	41.4	440	1	A7LA_MOUSE	Q9qy84	mus musculu
32	41	41.4	870	1	PLSB_XYLFA	Q9pej7	xylella fas
33	41	41.4	885	1	PLSB_XANAC	Q8pes0	xanthomonas
34	41	41.4	886	1	PLSB_XANCP	Q8p3e3	xanthomonas
35	41	41.4	1228	1	NARG_BACSU	P42175	bacillus su
36	41	41.4	1245	1	NARZ_ECOLI	P19319	escherichia
37	41	41.4	1246	1	NARG_ECOLI	P09152	escherichia
38	41	41.4	2214	1	POLG_CXA24	P36290	c genome po
39	40.5	40.9	191	1	AMEX_HUMAN	Q99217	homo sapien
40	40.5	40.9	565	1	ES10_RAT	P16303	rattus norv
41	40	40.4	246	1	IF28_HUMAN	Q96dx8	homo sapien
42	40	40.4	323	1	HEMZ_HAEIN	P43868	haemophilus
43	40	40.4	488	1	STCW_EMENI	Q00730	emericella
44	40	40.4	510	1	C304_DROME	Q9vg17	drosophila
45	40	40.4	600	1	DNLI_ACIAM	Q02093	acidianus a

ALIGNMENTS

RESULT 1
CAN2_MOUSE
ID CAN2_MOUSE STANDARD; PRT; 700 AA.
AC O08529; O35518; O54843;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
DE (Millimolar-calpain) (80 kDa M-calpain subunit) (CALP80).
GN CAPN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=97480729; PubMed=9339374;
RA Dear T.N., Matena K., Vingron M., Boehm T.;
RT "A new subfamily of vertebrate calpains lacking a calmodulin-like
RT domain: implications for calpain regulation and evolution.";
RL Genomics 45:175-184(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Ozaki Y.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=CNS;
RA Glass J.D., Nash N.R., Dry I., Culver D., Wesselingh S.;
RT "Cloning of m-calpain from mouse nervous system.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal tranduction (By similarity).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC Arg-|-Xaa with Leu or Val as the P2 residue.
CC -!- COFACTOR: Binds 3 calcium ions.
CC -!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
CC of calcium and inhibited by calpastatin.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding.
CC -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC ---

```
CC   ---------------------------------------------------------------------
DR   EMBL; Y10139; CAA71227.1; -.
DR   EMBL; D38117; BAA22964.1; -.
DR   EMBL; AF015038; AAB94029.1; -.
DR   HSSP; Q07009; 1DF0.
DR   MEROPS; C02.002; -.
DR   MGD; MGI:88264; Capn2.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR001300; Protease_C2.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF01067; Calpain_III; 1.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF00648; Peptidase_C2; 1.
DR   PRINTS; PR00704; CALPAIN.
DR   SMART; SM00720; calpain_III; 1.
DR   SMART; SM00230; CysPc; 1.
DR   SMART; SM00054; EFh; 2.
DR   PROSITE; PS00018; EF_HAND; 2.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW   Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family.
FT   PROPEP        1     19       ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT   CHAIN        20    700       CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT   DOMAIN       20    355       CALPAIN.
FT   DOMAIN      356    514       DOMAIN III.
FT   DOMAIN      515    529       LINKER.
FT   DOMAIN      530    700       DOMAIN IV.
FT   CA_BIND     541    552       EF-HAND 1.
FT   CA_BIND     585    596       EF-HAND 2.
FT   CA_BIND     615    626       EF-HAND 3.
FT   DOMAIN      650    661       ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT   DOMAIN      680    691       ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT   ACT_SITE    105    105       BY SIMILARITY.
FT   ACT_SITE    262    262       BY SIMILARITY.
FT   ACT_SITE    286    286       BY SIMILARITY.
FT   CONFLICT    194    194       A -> T (IN REF. 1).
FT   CONFLICT    212    212       A -> G (IN REF. 2).
FT   CONFLICT    402    402       E -> G (IN REF. 1).
SQ   SEQUENCE   700 AA;  79871 MW;  682146B290968316 CRC64;

  Query Match             56.6%;  Score 56;  DB 1;  Length 700;
  Best Local Similarity   66.7%;  Pred. No. 0.76;
  Matches   10;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || |||::|||| 
Db         52 PALPSSLGYKELGPY 66


RESULT 2
CAN2_RAT
ID   CAN2_RAT       STANDARD;      PRT;   700 AA.
AC   Q07009;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
DE   (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
DE   (Millimolar-calpain).
GN   CAPN2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94032492; PubMed=8218419;
RA   Deluca C.I., Davies P.L., Samis J.A., Elce J.S.;
RT   "Molecular cloning and bacterial expression of cDNA for rat calpain
RT   II 80 kDa subunit.";
RL   Biochim. Biophys. Acta 1216:81-93(1993).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=21240297; PubMed=11342050;
RA   Moldoveanu T., Hosfield C.M., Jia Z., Elce J.S., Davies P.L.;
RT   "Ca(2+)-induced structural changes in rat m-calpain revealed by
RT   partial proteolysis.";
RL   Biochim. Biophys. Acta 1545:245-254(2001).
RN   [3]
RP   MUTAGENESIS OF LYS-230; LYS-234 AND GLU-504.
RX   MEDLINE=21269273; PubMed=11102442;
RA   Hosfield C.M., Moldoveanu T., Davies P.L., Elce J.S., Jia Z.;
RT   "Calpain mutants with increased Ca2+ sensitivity and implications for
RT   the role of the C(2)-like domain.";
RL   J. Biol. Chem. 276:7404-7407(2001).
RN   [4]
RP   MUTAGENESIS OF CYS-105; HIS-262; ASN-286 AND TRP-288.
RX   MEDLINE=95361909; PubMed=7635186;
RA   Arthur J.S., Gauthier S., Elce J.S.;
RT   "Active site residues in m-calpain: identification by site-directed
RT   mutagenesis.";
RL   FEBS Lett. 368:397-400(1995).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX   MEDLINE=20069318; PubMed=10601010;
RA   Hosfield C.M., Elce J.S., Davies P.L., Jia Z.;
RT   "Crystal structure of calpain reveals the structural basis for
RT   Ca(2+)-dependent protease activity and a novel mode of enzyme
RT   activation.";
RL   EMBO J. 18:6880-6889(1999).
CC   -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC       catalyze limited proteolysis of substrates involved in
CC       cytoskeletal remodelling and signal tranduction.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC       Arg-|-Xaa with Leu or Val as the P2 residue.
CC   -!- COFACTOR: Binds 3 calcium ions.
CC   -!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
CC       of calcium and inhibited by calpastatin.
CC   -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC       (regulatory) subunit.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC       membrane upon Ca++ binding.
CC   -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09120; AAA16327.1; -.
DR   PIR; S38361; S38361.
DR   PDB; 1DF0; 26-NOV-01.
DR   MEROPS; C02.002; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR001300; Protease_C2.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF01067; Calpain_III; 1.
DR   Pfam; PF00036; efhand; 3.
DR   Pfam; PF00648; Peptidase_C2; 1.
DR   PRINTS; PR00704; CALPAIN.
DR   SMART; SM00720; calpain_III; 1.
DR   SMART; SM00230; CysPc; 1.
DR   SMART; SM00054; EFh; 2.
DR   PROSITE; PS00018; EF_HAND; 2.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW   Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family;
KW   3D-structure.
FT   PROPEP        1     19       ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT   CHAIN        20    700       CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
```

```
FT   DOMAIN       20    355       CALPAIN.
FT   DOMAIN      356    514       DOMAIN III.
FT   DOMAIN      515    529       LINKER.
FT   DOMAIN      530    700       DOMAIN IV.
FT   CA_BIND     541    552       EF-HAND 1.
FT   CA_BIND     585    596       EF-HAND 2.
FT   CA_BIND     615    626       EF-HAND 3.
FT   DOMAIN      650    661       ANCESTRAL CALCIUM SITE 4.
FT   DOMAIN      680    691       ANCESTRAL CALCIUM SITE 5.
FT   ACT_SITE    105    105
FT   ACT_SITE    262    262
FT   ACT_SITE    286    286
FT   MUTAGEN     105    105       C->S: LOSS OF ACTIVITY.
FT   MUTAGEN     226    226       K->S: DECREASE OF 12% OF THE ACTIVITY.
FT   MUTAGEN     230    230       K->S: NO EFFECT.
FT   MUTAGEN     230    230       K->E: DECREASE OF 84% OF THE ACTIVITY.
FT   MUTAGEN     234    234       K->E: DECREASE OF 85% OF THE ACTIVITY.
FT   MUTAGEN     234    234       K->S: DECREASE OF 20% OF THE ACTIVITY.
FT   MUTAGEN     262    262       H->A: LOSS OF ACTIVITY.
FT   MUTAGEN     286    286       N->A: LOSS OF ACTIVITY.
FT   MUTAGEN     288    288       W->Y: DECREASE OF 95% OF THE ACTIVITY.
FT   MUTAGEN     504    504       E->S: DECREASE OF 10% OF THE ACTIVITY.
SQ   SEQUENCE   700 AA;  79919 MW;  296B0DC3BEEF5B90 CRC64;

  Query Match           52.5%;  Score 52;  DB 1;  Length 700;
  Best Local Similarity  60.0%;  Pred. No. 3;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PALPSSLGFKELGPY 66


RESULT 3
HBPA_WHEAT
ID   HBPA_WHEAT     STANDARD;      PRT;   349 AA.
AC   P23922;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription factor HBP-1a (Histone-specific transcription factor
DE   HBP1).
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91224097; PubMed=2026143;
RA   Tabata T., Nakayama T., Mikami K., Iwabuchi M.;
RT   "HBP-1a and HBP-1b: leucine zipper-type transcription factors of
RT   wheat.";
RL   EMBO J. 10:1459-1467(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89368924; PubMed=2772648;
RA   Tabata T., Takase H., Takayama S., Mikami K., Nakatsuka A., Kawata T.,
RA   Nakayama T., Iwabuchi M.;
RT   "A protein that binds to a cis-acting element of wheat histone genes
RT   has a leucine zipper motif.";
RL   Science 245:965-967(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Horoshirikomugi;
RX   MEDLINE=96027929; PubMed=7476857;
RA   Mikami K., Katsura M., Ito T., Okada K., Shimura Y., Iwabuchi M.;
RT   "Developmental and tissue-specific regulation of the gene for the
RT   wheat basic/leucine zipper protein HBP-1a(17) in transgenic
RT   Arabidopsis plants.";
RL   Mol. Gen. Genet. 248:573-582(1995).
CC   -!- FUNCTION: BIND TO THE HEXAMER MOTIF 5'-ACGTCA-3' OF HISTONE
CC       GENE PROMOTERS.
CC   -!- SUBUNIT: Binds DNA as a dimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the bZIP family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X56781; CAA40101.1; -.
DR   EMBL; M28704; AAA34293.1; -.
DR   EMBL; D38111; BAA07289.1; -.
DR   PIR; S77570; S77570.
DR   HSSP; P03069; 2DGC.
DR   TRANSFAC; T00354; -.
DR   TRANSFAC; T00937; -.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   PROSITE; PS50217; BZIP; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT   DNA_BIND     254    273       BASIC MOTIF.
FT   DOMAIN       280    294       LEUCINE-ZIPPER.
SQ   SEQUENCE   349 AA;  36743 MW;  51FCC8E039DF7F7A CRC64;

  Query Match           51.5%;  Score 51;  DB 1;  Length 349;
  Best Local Similarity  46.7%;  Pred. No. 2.2;
  Matches    7;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |||| : ||  : |:
Db         34 PEWPGFQGYPAMPPH 48


RESULT 4
CAN2_HUMAN
ID   CAN2_HUMAN     STANDARD;      PRT;   700 AA.
AC   P17655; Q16738; Q8WU26; Q9HBB1;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
DE   (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
DE   (Millimolar-calpain) (Calpain large polypeptide L2).
GN   CAPN2 OR CANPL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89166474; PubMed=2852952;
RA   Imajoh S., Aoki K., Ohno S., Emori Y., Kawasaki H., Sugihara H.,
RA   Suzuki K.;
RT   "Molecular cloning of the cDNA for the large subunit of the
RT   high-Ca2+-requiring form of human Ca2+-activated neutral protease.";
RL   Biochemistry 27:8122-8128(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Astrocytoma;
RX   MEDLINE=20403900; PubMed=10944468;
RA   Ye Z., Connor J.R.;
RT   "cDNA cloning by amplification of circularized first strand cDNAs
RT   reveals non-IRE-regulated iron-responsive mRNAs.";
RL   Biochem. Biophys. Res. Commun. 275:223-227(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
```

```
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 1-79 FROM N.A.
RC   TISSUE=Lymph node;
RX   MEDLINE=89197947; PubMed=2539381;
RA   Hata A., Ohno S., Akita Y., Suzuki K.;
RT   "Tandemly reiterated negative enhancer-like elements regulate
RT   transcription of a human gene for the large subunit of calcium-
RT   dependent protease.";
RL   J. Biol. Chem. 264:6404-6411(1989).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX   MEDLINE=20105516; PubMed=10639123;
RA   Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H.,
RA   Nakagawa K., Irie A., Sorimachi H., Bourenkow G., Bartunik H.,
RA   Suzuki K., Bode W.;
RT   "The crystal structure of calcium-free human m-calpain suggests an
RT   electrostatic switch mechanism for activation by calcium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
CC   -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC       catalyze limited proteolysis of substrates involved in
CC       cytoskeletal remodelling and signal tranduction.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC       Arg-|-Xaa with Leu or Val as the P2 residue.
CC   -!- COFACTOR: Binds 3 calcium ions.
CC   -!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
CC       of calcium and inhibited by calpastatin.
CC   -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC       (regulatory) subunit.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC       membrane upon Ca++ binding.
CC   -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23254; AAA35645.1; -.
DR   EMBL; AF261089; AAF99682.1; -.
DR   EMBL; BC021303; AAH21303.1; -.
DR   EMBL; J04700; AAA52760.1; -.
DR   PIR; S10590; CIHUH2.
DR   PDB; 1KFU; 07-DEC-01.
DR   PDB; 1KFX; 07-DEC-01.
DR   MEROPS; C02.002; -.
DR   Genew; HGNC:1479; CAPN2.
DR   MIM; 114230; -.
DR   GO; GO:0008234; F:cysteine-type peptidase activity; TAS.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR001300; Protease_C2.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF01067; Calpain_III; 1.
DR   Pfam; PF00036; efhand; 3.
DR   Pfam; PF00648; Peptidase_C2; 1.
DR   PRINTS; PR00704; CALPAIN.
DR   SMART; SM00720; calpain_III; 1.
DR   SMART; SM00230; CysPc; 1.
DR   PROSITE; PS00018; EF_HAND; 2.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW   Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family;
KW   3D-structure; Polymorphism.
FT   PROPEP          1     19       ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT   CHAIN          20    700       CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT   DOMAIN         27    355       CALPAIN.
FT   DOMAIN        356    514       DOMAIN III.
FT   DOMAIN        515    529       LINKER.
FT   DOMAIN        530    700       DOMAIN IV.
FT   CA_BIND       541    552       EF-HAND 1.
FT   CA_BIND       585    596       EF-HAND 2.
FT   CA_BIND       615    626       EF-HAND 3.
FT   DOMAIN        650    661       ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT   DOMAIN        680    691       ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT   ACT_SITE      105    105       BY SIMILARITY.
FT   ACT_SITE      262    262       BY SIMILARITY.
FT   ACT_SITE      286    286       BY SIMILARITY.
FT   VARIANT        22     22       E -> D (IN dbSNP:25655).
FT                                  /FTId=VAR_014435.
FT   VARIANT       568    568       K -> Q (IN dbSNP:17599).
FT                                  /FTId=VAR_014436.
FT   CONFLICT       68     68       S -> G (IN REF. 4).
FT   CONFLICT       73     74       IE -> MR (IN REF. 1).
FT   CONFLICT      256    256       Q -> K (IN REF. 2).
FT   CONFLICT      300    300       N -> S (IN REF. 2).
FT   CONFLICT      534    534       V -> F (IN REF. 3).
FT   HELIX           4     16
FT   TURN           17     19
FT   TURN           22     23
FT   STRAND         25     26
FT   HELIX          27     29
FT   HELIX          32     42
FT   TURN           43     43
FT   TURN           55     58
FT   TURN           69     70
FT   STRAND         73     76
FT   HELIX          78     81
FT   STRAND         86     86
FT   STRAND         97     98
FT   TURN          101    102
FT   HELIX         105    113
FT   TURN          114    115
FT   HELIX         118    121
FT   TURN          122    124
FT   TURN          136    137
FT   STRAND        138    144
FT   STRAND        149    155
FT   STRAND        158    161
FT   TURN          162    163
FT   STRAND        164    165
FT   STRAND        169    170
FT   STRAND        175    175
FT   HELIX         177    187
FT   TURN          188    189
FT   HELIX         192    194
FT   TURN          196    197
FT   HELIX         200    206
FT   TURN          207    208
FT   STRAND        211    216
FT   TURN          217    218
FT   TURN          222    223
FT   HELIX         224    232
```

```
FT   TURN        233    235
FT   STRAND      237    241
FT   TURN        253    254
FT   STRAND      264    272
FT   TURN        275    276
FT   STRAND      279    285
FT   TURN        287    288
FT   TURN        303    304
FT   HELIX       310    313
FT   TURN        314    317
FT   STRAND      327    327
FT   HELIX       328    332
FT   STRAND      336    341
FT   HELIX       344    346
FT   TURN        351    352
FT   STRAND      357    357
FT   STRAND      363    365
FT   TURN        367    370
FT   TURN        375    376
FT   TURN        378    380
FT   HELIX       381    383

  Query Match          .   51.5%;  Score 51;  DB 1;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 4.2;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 5
CANX_CHICK
ID   CANX_CHICK     STANDARD;      PRT;   705 AA.
AC   P00789;
DT   21-JUL-1986 (Rel. 01, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Calpain, large [catalytic] subunit (EC 3.4.22.17) (Calcium-
DE   activated neutral proteinase) (CANP) (Mu/M-type).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85061606; PubMed=6095110;
RA   Ohno S., Emori Y., Imajoh S., Kawasaki H., Kisaragi M., Suzuki K.;
RT   "Evolutionary origin of a calcium-dependent protease by fusion of
RT   genes for a thiol protease and a calcium-binding protein?";
RL   Nature 312:566-570(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86082358; PubMed=3000828;
RA   Emori Y., Ohno S., Tobita M., Suzuki K.;
RT   "Gene structure of calcium-dependent protease retains the ancestral
RT   organization of the calcium-binding protein gene.";
RL   FEBS Lett. 194:249-252(1986).
RN   [3]
RP   CALCIUM-BINDING DATA.
RX   MEDLINE=87279982; PubMed=3038855;
RA   Minami Y., Emori Y., Kawasaki H., Suzuki K.;
RT   "E-F hand structure-domain of calcium-activated neutral protease
RT   (CANP) can bind Ca2+ ions.";
RL   J. Biochem. 101:889-895(1987).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=95260862; PubMed=7742367;
RA   Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
RA   Suzuki K.;
RT   "Identification of a third ubiquitous calpain species -- chicken
RT   muscle expresses four distinct calpains.";
RL   Biochim. Biophys. Acta 1261:381-393(1995).
CC   -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC       catalyze limited proteolysis of substrates involved in
CC       cytoskeletal remodelling and signal tranduction.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC       Arg-|-Xaa with Leu or Val as the P2 residue.
CC   -!- COFACTOR: Binds 3 calcium ions.
CC   -!- ENZYME REGULATION: Activated by micromolar concentrations of
CC       calcium and inhibited by calpastatin.
CC   -!- SUBUNIT: Heterodimer of large (catalytic) and a small (regulatory)
CC       subunit.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC       membrane upon Ca++ binding (By similarity).
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC   -!- CAUTION: THIS PROTEIN WAS PREVIOUSLY THOUGHT TO BE M-CALPAIN BUT
CC       HAS SINCE BEEN FOUND TO BE AN INTERMEDIATE FORM BETWEEN THE M AND
CC       MU TYPES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X01415; CAA25658.1; -.
DR   PIR; A00979; CICHH.
DR   HSSP; Q07009; 1DF0.
DR   MEROPS; C02.003; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR001300; Protease_C2.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF01067; Calpain_III; 1.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF00648; Peptidase_C2; 1.
DR   PRINTS; PR00704; CALPAIN.
DR   SMART; SM00720; calpain_III; 1.
DR   SMART; SM00230; CysPc; 1.
DR   SMART; SM00054; EFh; 2.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW   Hydrolase; Thiol protease; Calcium-binding; Multigene family.
FT   DOMAIN       23    358       CALPAIN.
FT   DOMAIN      359    517       DOMAIN III.
FT   DOMAIN      518    533       LINKER.
FT   DOMAIN      534    704       DOMAIN IV.
FT   CA_BIND     545    556       EF-HAND 1.
FT   CA_BIND     589    600       EF-HAND 2.
FT   CA_BIND     619    630       EF-HAND 3.
FT   DOMAIN      654    665       ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT   DOMAIN      684    695       ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT   ACT_SITE    108    108       BY SIMILARITY.
FT   ACT_SITE    265    265       BY SIMILARITY.
FT   ACT_SITE    289    289       BY SIMILARITY.
SQ   SEQUENCE   705 AA;  80351 MW;  ABCDDC56298E48AA CRC64;

  Query Match          49.5%;  Score 49;  DB 1;  Length 705;
  Best Local Similarity   53.3%;  Pred. No. 8.4;
  Matches    8;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  |: ||:::|||| 
Db         55 PAGPTALGFKELGPY 69


RESULT 6
DNLI_SULTO
```

```
ID   DNLI_SULTO     STANDARD;      PRT;   600 AA.
AC   Q976G4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Thermostable DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide
DE   synthase [ATP]).
GN   LIG OR ST0223.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   MEDLINE=21456156; PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
CC   -!- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
CC       RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +
CC       {deoxyribonucleotide}(M) = AMP + diphosphate +
CC       {deoxyribonucleotide}(N+M).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000981; BAB65183.1; -.
DR   HAMAP; MF_00407; -; 1.
DR   InterPro; IPR000977; DNA_ligase.
DR   Pfam; PF01068; DNA_ligase; 1.
DR   Pfam; PF04679; DNA_ligase_A_C; 1.
DR   Pfam; PF04675; DNA_ligase_A_N; 1.
DR   TIGRFAMs; TIGR00574; dnl1; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR   PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR   PROSITE; PS50160; DNA_LIGASE_A3; 1.
KW   DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW   ATP-binding; Complete proteome.
FT   BINDING     260    260       AMP (BY SIMILARITY).
SQ   SEQUENCE   600 AA;  68025 MW;  9D85DA4458000539 CRC64;

  Query Match             48.5%;  Score 48;  DB 1;  Length 600;
  Best Local Similarity   63.6%;  Pred. No. 10;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 WPSYLGYEKLG 13
              || :||| :||
Db         50 WPDFLGYPELG 60


RESULT 7
DNLI_SULSH
ID   DNLI_SULSH     STANDARD;      PRT;   601 AA.
AC   Q9P9K9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Thermostable DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide
DE   synthase [ATP]).
GN   LIG.
OS   Sulfolobus shibatae.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2286;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lai X., Shao H., Huang L.;
RT   "A thermophilic DNA ligase from Sulfolobus shibatae.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
CC       RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +
CC       {deoxyribonucleotide}(M) = AMP + diphosphate +
CC       {deoxyribonucleotide}(N+M).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF242877; AAF61267.1; -.
DR   HAMAP; MF_00407; -; 1.
DR   InterPro; IPR000977; DNA_ligase.
DR   Pfam; PF01068; DNA_ligase; 1.
DR   Pfam; PF04679; DNA_ligase_A_C; 1.
DR   Pfam; PF04675; DNA_ligase_A_N; 1.
DR   TIGRFAMs; TIGR00574; dnl1; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR   PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR   PROSITE; PS50160; DNA_LIGASE_A3; 1.
KW   DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW   ATP-binding.
FT   BINDING     260    260       AMP (BY SIMILARITY).
SQ   SEQUENCE   601 AA;  67646 MW;  C029AC84AB39AFD4 CRC64;

  Query Match             48.5%;  Score 48;  DB 1;  Length 601;
  Best Local Similarity   63.6%;  Pred. No. 10;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 WPSYLGYEKLG 13
              || :||| :||
Db         50 WPDFLGYPELG 60


RESULT 8
DNLI_SULSO
ID   DNLI_SULSO     STANDARD;      PRT;   601 AA.
AC   Q980T8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Thermostable DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide
DE   synthase [ATP]).
GN   LIG OR SSO0189.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
```

```
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC   -!- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
CC       RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +
CC       {deoxyribonucleotide}(M) = AMP + diphosphate +
CC       {deoxyribonucleotide}(N+M).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006656; AAK40535.1; -.
DR   PIR; H90159; H90159.
DR   HAMAP; MF_00407; -; 1.
DR   InterPro; IPR000977; DNA_ligase.
DR   Pfam; PF01068; DNA_ligase; 1.
DR   Pfam; PF04679; DNA_ligase_A_C; 1.
DR   Pfam; PF04675; DNA_ligase_A_N; 1.
DR   TIGRFAMs; TIGR00574; dnl1; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR   PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR   PROSITE; PS50160; DNA_LIGASE_A3; 1.
KW   DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW   ATP-binding; Complete proteome.
FT   BINDING     260    260       AMP (BY SIMILARITY).
SQ   SEQUENCE   601 AA;  67732 MW;  DA6814F4A6F0546E CRC64;

  Query Match             48.5%;  Score 48;  DB 1;  Length 601;
  Best Local Similarity   63.6%;  Pred. No. 10;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 WPSYLGYEKLG 13
              || :||| :||
Db         50 WPDFLGYPELG 60


RESULT 9
CRK1_LEIME
ID   CRK1_LEIME     STANDARD;      PRT;   301 AA.
AC   Q06309;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Cell division protein kinase 2 homolog CRK1 (EC 2.7.1.-).
GN   CRK1.
OS   Leishmania mexicana.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ssp. Mexicana;
RX   MEDLINE=94012652; PubMed=8407941;
RA   Mottram J.C., Kinnaird J.H., Shiels B.R., Tait A., Barry J.D.;
RT   "A novel CDC2-related protein kinase from Leishmania mexicana,
RT   LmmCRK1, is post-translationally regulated during the life cycle.";
RL   J. Biol. Chem. 268:21044-21052(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN SOME STAGE-SPECIFIC ROLE IN THE
CC       PROMASTIGOTE CELL CYCLE.
CC   -!- ENZYME REGULATION: PHOSPHORYLATION AT THR-15 OR TYR-16 INACTIVATES
CC       THE ENZYME, WHILE PHOSPHORYLATION AT THR-160 ACTIVATES IT (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: FORMS A STABLE BUT NON-COVALENT COMPLEX WITH A REGULATORY
CC       SUBUNIT AND WITH A CYCLIN (BY SIMILARITY).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL LIFE CYCLE STAGES,
CC       PROMASTIGOTE, METACYCLIC AND AMASTIGOTE FORMS BUT IS FOUND
CC       IN THE ACTIVE FORM ONLY IN THE PROMASTIGOTE STAGE.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       CDC2/CDKX SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60385; CAA42936.1; -.
DR   PIR; A48041; A48041.
DR   HSSP; P24941; 1BUH.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation.
FT   DOMAIN        5    297       PROTEIN KINASE.
FT   NP_BIND      11     19       ATP (BY SIMILARITY).
FT   BINDING      34     34       ATP (BY SIMILARITY).
FT   ACT_SITE    127    127       BY SIMILARITY.
FT   MOD_RES      15     15       PHOSPHORYLATION.
FT   MOD_RES      16     16       PHOSPHORYLATION.
FT   MOD_RES     160    160       PHOSPHORYLATION (BY CAK) (BY SIMILARITY).
SQ   SEQUENCE   301 AA;  34473 MW;  58EB39D006D88461 CRC64;

  Query Match             47.5%;  Score 47;  DB 1;  Length 301;
  Best Local Similarity   58.8%;  Pred. No. 7.4;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    4;  Gaps    1;

Qy          1 PEWPSYL----GYEKLG 13
              ||| : |    ||||||
Db        251 PEWSNVLGSVPGYEKLG 267


RESULT 10
GALM_HAEIN
ID   GALM_HAEIN     STANDARD;      PRT;   340 AA.
AC   P31765;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Aldose 1-epimerase (EC 5.1.3.3) (Mutarotase).
GN   GALM OR MRO OR HI0818.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
```

```
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE OF 1-116 FROM N.A.
RC   STRAIN=RM 7004 / Serotype B;
RX   MEDLINE=93125127; PubMed=1282642;
RA   Maskell D.J., Szabo M.J., Deadman M.E., Moxon E.R.;
RT   "The gal locus from Haemophilus influenzae: cloning, sequencing and
RT   the use of gal mutants to study lipopolysaccharide.";
RL   Mol. Microbiol. 6:3051-3063(1992).
CC   -!- FUNCTION: Mutarotase converts alpha-aldose to the beta-anomer. It
CC       is active on D-glucose, L-arabinose, D-xylose, D-galactose,
CC       maltose and lactose (By similarity).
CC   -!- CATALYTIC ACTIVITY: Alpha-D-glucose = beta-D-glucose.
CC   -!- PATHWAY: Hexose metabolism.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE ALDOSE EPIMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32764; AAC22477.1; -.
DR   EMBL; X65934; CAA46732.1; -.
DR   PIR; C64096; C64096.
DR   TIGR; HI0818; -.
DR   InterPro; IPR001823; Aldl_epimerase.
DR   Pfam; PF01263; Aldose_epim; 1.
DR   PROSITE; PS00545; ALDOSE_1_EPIMERASE; 1.
KW   Isomerase; Complete proteome.
FT   ACT_SITE    172    172       POTENTIAL.
FT   VARIANT      42     42       V -> I (IN STRAIN RM 7004).
FT   VARIANT      46     46       D -> G (IN STRAIN RM 7004).
FT   VARIANT      58     58       D -> E (IN STRAIN RM 7004).
FT   VARIANT     113    113       N -> K (IN STRAIN RM 7004).
SQ   SEQUENCE   340 AA;  38102 MW;  4ECAE4FB8C498C7A CRC64;

  Query Match             46.5%;  Score 46;  DB 1;  Length 340;
  Best Local Similarity   50.0%;  Pred. No. 12;
  Matches    8;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPYY 16
              ||| :| | :| |  |
Db        316 PEWQNYGGIQKAGGRY 331


RESULT 11
FETA_RAT
ID   FETA_RAT       STANDARD;      PRT;   611 AA.
AC   P02773; Q63032;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Alpha-fetoprotein precursor (Alpha-fetoglobulin) (Alpha-1-
DE   fetoprotein).
GN   AFP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-106 FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=85215621; PubMed=2582363;
RA   Turcotte B., Guertin M., Chevrette M., Belanger L.;
RT   "Rat alpha 1-fetoprotein messenger RNA: 5'-end sequence and
RT   glucocorticoid-suppressed liver transcription in an improved nuclear'
RT   run-off assay.";
RL   Nucleic Acids Res. 13:2387-2398(1985).
RN   [2]
RP   SEQUENCE OF 91-611 FROM N.A.
RX   MEDLINE=81273091; PubMed=6167988;
RA   Jagodzinski L.L., Sargent T.D., Yang M., Glackin C., Bonner J.;
RT   "Sequence homology between RNAs encoding rat alpha-fetoprotein and
RT   rat serum albumin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 78:3521-3525(1981).
RN   [3]
RP   SEQUENCE OF 1-51 FROM N.A.
RX   MEDLINE=92119318; PubMed=1722723;
RA   Buzard G.S., Locker J.;
RT   "The transcription control region of the rat alpha-fetoprotein gene.
RT   DNA sequence and homology studies.";
RL   DNA Seq. 1:33-48(1990).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=87146445; PubMed=2434929;
RA   Chevrette M., Guertin M., Turcotte B., Belanger L.;
RT   "The rat alpha 1-fetoprotein gene: characterization of the
RT   5'-flanking region and tandem organization with the albumin gene.";
RL   Nucleic Acids Res. 15:1338-1338(1987).
RN   [5]
RP   SEQUENCE OF 1-13 FROM N.A.
RX   MEDLINE=87308273; PubMed=2442163;
RA   Nahon J.L., Danan J.L., Poiret M., Tratner I., Jose-Estanyol M.,
RA   Sala-Trepat J.M.;
RT   "The rat alpha-fetoprotein and albumin genes. Transcriptional control
RT   and comparison of the sequence organization and promoter region.";
RL   J. Biol. Chem. 262:12479-12487(1987).
RN   [6]
RP   SEQUENCE OF 556-611 FROM N.A.
RX   MEDLINE=81006806; PubMed=6157681;
RA   Liao W.S.L., Hamilton R.W., Taylor J.M.;
RT   "Amino acid sequence homology between rat alpha-fetoprotein and
RT   albumin at the COOH-terminal regions.";
RL   J. Biol. Chem. 255:8046-8049(1980).
CC   -!- FUNCTION: Binds estrogens, fatty acids and metals.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Plasma.
CC   -!- PTM: SULFATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC   -!- SIMILARITY: Contains 3 albumin domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02361; CAA26214.1; -.
DR   EMBL; V01254; CAA24567.1; -.
DR   EMBL; J00695; AAA40697.1; -.
DR   EMBL; X05093; CAA28744.1; -.
DR   EMBL; J02816; AAA40695.1; ALT_INIT.
DR   EMBL; J00694; AAA40694.1; -.
DR   PIR; A93561; FPRT.
DR   HSSP; P02768; 1E7B.
DR   InterPro; IPR000264; Serum_albumin.
DR   Pfam; PF00273; transport_prot; 3.
DR   PRINTS; PR00802; SERUMALBUMIN.
DR   ProDom; PD002486; Serum_albumin; 1.
DR   SMART; SM00103; ALBUMIN; 3.
DR   PROSITE; PS00212; ALBUMIN; 2.
KW   Glycoprotein; Sulfation; Repeat; Metal-binding; Signal.
FT   SIGNAL      1     24       PROBABLE.
FT   CHAIN      25    611       ALPHA-FETOPROTEIN.
FT   DOMAIN     26    207       ALBUMIN 1.
FT   DOMAIN    214    399       ALBUMIN 2.
FT   DOMAIN    406    597       ALBUMIN 3.
FT   METAL      28     28       COPPER (POTENTIAL).
FT   DISULFID  101    116       BY SIMILARITY.
FT   DISULFID  115    126       BY SIMILARITY.
```

```
FT   DISULFID    150    195       BY SIMILARITY.
FT   DISULFID    194    203       BY SIMILARITY.
FT   DISULFID    226    272       BY SIMILARITY.
FT   DISULFID    271    279       BY SIMILARITY.
FT   DISULFID    291    305       BY SIMILARITY.
FT   DISULFID    304    315       BY SIMILARITY.
FT   DISULFID    386    395       BY SIMILARITY.
FT   DISULFID    418    464       BY SIMILARITY.
FT   DISULFID    463    474       BY SIMILARITY.
FT   DISULFID    487    503       BY SIMILARITY.
FT   DISULFID    502    513       BY SIMILARITY.
FT   DISULFID    540    585       BY SIMILARITY.
FT   DISULFID    584    593       BY SIMILARITY.
FT   CARBOHYD    117    117       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    556    557       IN -> VG (IN REF. 6).
SQ   SEQUENCE   611 AA;  68386 MW;  7DC54624B7245C41 CRC64;

  Query Match           45.5%;  Score 45;  DB 1;  Length 611;
  Best Local Similarity  87.5%;  Pred. No. 29;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          9 YEKLGPYY 16
              |:||||||
Db        421 YQKLGPYY 428


RESULT 12
CAN2_CHICK
ID   CAN2_CHICK     STANDARD;      PRT;   700 AA.
AC   Q92178;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
DE   (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
DE   (Millimolar-calpain).
GN   CAPN2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=95260862; PubMed=7742367;
RA   Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
RA   Suzuki K.;
RT   "Identification of a third ubiquitous calpain species -- chicken
RT   muscle expresses four distinct calpains.";
RL   Biochim. Biophys. Acta 1261:381-393(1995).
CC   -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC       catalyze limited proteolysis of substrates involved in
CC       cytoskeletal remodelling and signal tranduction (By similarity).
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC       Arg-|-Xaa with Leu or Val as the P2 residue.
CC   -!- COFACTOR: Binds 3 calcium ions.
CC   -!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
CC       of calcium and inhibited by calpastatin.
CC   -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC       (regulatory) subunit.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC       membrane upon Ca++ binding.
CC   -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D38026; BAA07228.1; -.
DR   PIR; S57194; S57194.
DR   HSSP; Q07009; 1DF0.
DR   MEROPS; C02.002; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR001300; Protease_C2.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF01067; Calpain_III; 1.
DR   Pfam; PF00036; efhand; 3.
DR   Pfam; PF00648; Peptidase_C2; 1.
DR   PRINTS; PR00704; CALPAIN.
DR   SMART; SM00720; calpain_III; 1.
DR   SMART; SM00230; CysPc; 1.
DR   SMART; SM00054; EFh; 3.
DR   PROSITE; PS00018; EF_HAND; 2.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW   Hydrolase; Thiol protease; Calcium-binding; Multigene family.
FT   PROPEP        1     19       ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT   CHAIN        20    700       CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT   DOMAIN       20    355       CALPAIN.
FT   DOMAIN      356    514       DOMAIN III.
FT   DOMAIN      515    529       LINKER.
FT   DOMAIN      530    700       DOMAIN IV.
FT   CA_BIND     541    552       EF-HAND 1.
FT   CA_BIND     585    596       EF-HAND 2.
FT   CA_BIND     615    626       EF-HAND 3.
FT   DOMAIN      650    661       ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT   DOMAIN      680    691       ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT   ACT_SITE    105    105       BY SIMILARITY.
FT   ACT_SITE    262    262       BY SIMILARITY.
FT   ACT_SITE    286    286       BY SIMILARITY.
SQ   SEQUENCE   700 AA;  79228 MW;  C3AEDB39CCB56D3B CRC64;

  Query Match           45.5%;  Score 45;  DB 1;  Length 700;
  Best Local Similarity  57.1%;  Pred. No. 33;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGP 14
              |  |: ||| :|||
Db         52 PAGPAALGYRELGP 65


RESULT 13
SUF2_HUMAN
ID   SUF2_HUMAN     STANDARD;      PRT;   748 AA.
AC   Q9HAU4; Q9H260;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Smad ubiquitination regulatory factor 2 (EC 6.3.2.-) (Ubiquitin--
DE   protein ligase SMURF2) (Smad-specific E3 ubiquitin ligase 2)
DE   (hSMURF2).
GN   SMURF2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF PRO-251--VAL-284 AND
RP   GLY-297--LEU-330.
RX   PubMed=11163210;
RA   Kavsak P., Rasmussen R.K., Causing C.G., Bonni S., Zhu H.,
RA   Thomsen G.H., Wrana J.L.;
RT   "Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the
RT   TGF-beta receptor for degradation.";
RL   Mol. Cell 6:1365-1375(2000).
RN   [2]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF PRO-251--VAL-284 AND CYS-716.
RX   MEDLINE=20538422; PubMed=11016919;
```

```
RA   Lin X., Liang M., Feng X.-H.;
RT   "Smurf2 Is a ubiquitin E3 ligase mediating proteasome-dependent
RT   degradation of Smad2 in transforming growth factor-beta signaling.";
RL   J. Biol. Chem. 275:36818-36822(2000).
RN   [3]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF CYS-716.
RX   MEDLINE=21107656; PubMed=11158580;
RA   Zhang Y., Chang C., Gehling D.J., Hemmati-Brivanlou A., Derynck R.;
RT   "Regulation of Smad degradation and activity by Smurf2, an E3
RT   ubiquitin ligase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:974-979(2001).
CC   -!- FUNCTION: Interacts with SMAD1, SMAD2 and SMAD7 in order to
CC       trigger their ubiquitination and proteasome-dependent degradation.
CC       Enhances the inhibitory activity of SMAD7 and reduces the
CC       transcriptional activity of SMAD2. Coexpression of SMURF2 with
CC       SMAD1 results in considerable decrease in steady-state level of
CC       SMAD1 protein and a smaller decrease of SMAD2 level.
CC   -!- SUBUNIT: Interacts with SMAD1, SMAD2, SMAD3, SMAD6 and SMAD7 but
CC       not SMAD4.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Cytoplasmic in the presence of
CC       SMAD7.
CC   -!- TISSUE SPECIFICITY: Widely expressed.
CC   -!- DOMAIN: The second and third WW domains are responsible for
CC       interaction with R-SMAD (SMAD1, SMAD2 and SMAD3).
CC   -!- SIMILARITY: Contains 1 C2 domain.
CC   -!- SIMILARITY: Contains 3 WW domains.
CC   -!- SIMILARITY: Contains 1 HECT-type E3 ubiquitin-protein ligase
CC       domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF310676; AAG45422.1; -.
DR   EMBL; AF301463; AAG25641.1; -.
DR   EMBL; AY014180; AAG50421.1; -.
DR   HSSP; Q13526; 1PIN.
DR   MIM; 605532; -.
DR   GO; GO:0004842; F:ubiquitin-protein ligase activity; NAS.
DR   GO; GO:0016481; P:negative regulation of transcription; NAS.
DR   GO; GO:0017015; P:regulation of TGFbeta receptor signaling pa. . .; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR000569; HECT_domain.
DR   InterPro; IPR001202; WW_Rsp5_WWP.
DR   Pfam; PF00168; C2; 1.
DR   Pfam; PF00632; HECT; 1.
DR   Pfam; PF00397; WW; 3.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00119; HECTc; 1.
DR   SMART; SM00456; WW; 3.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
DR   PROSITE; PS50237; HECT; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; 1.
DR   PROSITE; PS50020; WW_DOMAIN_2; 3.
KW   Ubl conjugation pathway; Ligase; Repeat; Nuclear protein.
FT   DOMAIN        1     98       C2 DOMAIN.
FT   DOMAIN      157    190       WW 1.
FT   DOMAIN      251    284       WW 2.
FT   DOMAIN      297    330       WW 3.
FT   DOMAIN      414    748       HECT.
FT   MUTAGEN     251    284       MISSING: ABOLISHES INTERACTION WITH
FT                                SMAD2 AND SMAD7.
FT   MUTAGEN     297    330       MISSING: ABOLISHES INTERACTION WITH
FT                                SMAD7.
FT   MUTAGEN     716    716       C->A: LOSS OF ABILITY TO UBIQUITINATE
FT                                SMAD2.
FT   MUTAGEN     716    716       C->G: LOSS OF ACTIVITY (LOSS OF ABILITY
FT                                TO UBIQUITINATE SMAD1 AND SMAD2 AND NO
FT                                DOWN-REGULATION OF SMAD1 AND SMAD2
FT                                PROTEIN LEVELS).
FT   CONFLICT      6      6       G -> R (IN REF. 2).
SQ   SEQUENCE   748 AA;  86195 MW;  3042B443A3755762 CRC64;

  Query Match             45.5%;  Score 45;  DB 1;  Length 748;
  Best Local Similarity   53.3%;  Pred. No. 35;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 EWPSYLGYEKLGPYY 16
              ||   | :| | |||
Db        440 EWLYLLSHEMLNPYY 454


RESULT 14
CB2_MALDO
ID   CB2_MALDO      STANDARD;      PRT;   268 AA.
AC   P15773;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Chlorophyll A-B binding protein AB10, chloroplast precursor (LHCII
DE   type I CAB-AB10) (LHCP).
OS   Malus domestica (Apple) (Malus sylvestris).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Rosales; Rosaceae; Maloideae; Malus.
OX   NCBI_TaxID=3750;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Golden Delicious; TISSUE=Leaf;
RX   MEDLINE=90175017; PubMed=2408025;
RA   Chen H., Korban S.S., Buetow D.E.;
RT   "Nucleotide sequence of an apple nuclear gene encoding a light-
RT   harvesting chlorophyll a/b binding polypeptide of photosystem II.";
RL   Nucleic Acids Res. 18:679-679(1990).
CC   -!- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
CC       RECEPTOR, IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
CC       WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
CC       EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
CC       GRANAL MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
CC       OF ITS THREONINE RESIDUES; BOTH ARE BELIEVED TO MEDIATE THE
CC       DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
CC   -!- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND
CC       CHLOROPHYLL A-B BINDING PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17697; CAA35690.1; -.
DR   PIR; S08229; S08229.
DR   InterPro; IPR001344; Chloro_ABbind.
DR   Pfam; PF00504; chloroa_b-bind; 1.
DR   ProDom; PD000275; Chloro_ABbind; 1.
KW   Chlorophyll; Photosynthesis; Photosystem I; Photosystem II;
KW   Thylakoid; Membrane; Chloroplast; Transit peptide; Multigene family;
KW   Transmembrane; Phosphorylation.
FT   TRANSIT       1     40       CHLOROPLAST (PROBABLE).
FT   CHAIN        41    268       CHLOROPHYLL A-B BINDING PROTEIN AB10.
FT   TRANSMEM    103    122       POTENTIAL.
FT   TRANSMEM    155    175       POTENTIAL.
FT   TRANSMEM    222    237       POTENTIAL.
SQ   SEQUENCE   268 AA;  29211 MW;  1CAA091F7AEC6303 CRC64;

  Query Match             44.4%;  Score 44;  DB 1;  Length 268;
  Best Local Similarity   64.3%;  Pred. No. 18;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;
```

```
Qy          2 EWPSYLGYEKLGPY 15
              ||||||  |  | |
Db         68 EWPSYLTGEFPGDY 81


RESULT 15
LEU2_PYRAE
ID   LEU2_PYRAE     STANDARD;      PRT;   415 AA.
AC   Q8ZW41;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
DE   (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN   LEUC OR PAE1984.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   MEDLINE=21664397; PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC   -!- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC       and 3-isopropylmalate, via the formation of 2-isopropylmaleate.
CC   -!- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmaleate +
CC       H(2)O.
CC   -!- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-
CC       isopropylmalate.
CC   -!- COFACTOR: Binds 1 4Fe-4S cluster per subunit (By similarity).
CC   -!- PATHWAY: Leucine biosynthesis; second step.
CC   -!- SUBUNIT: Heterodimer of leuC and leuD (By similarity).
CC   -!- SIMILARITY: Belongs to the aconitase/IPM isomerase family. LeuC 2
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009850; AAL63861.1; -.
DR   HAMAP; MF_01027; -; 1.
DR   InterPro; IPR001030; Aconitase_N.
DR   InterPro; IPR006251; Cis-H_aconitase.
DR   Pfam; PF00330; aconitase; 1.
DR   PRINTS; PR00415; ACONITASE.
DR   ProDom; PD000511; Aconitase_N; 1.
DR   TIGRFAMs; TIGR01343; hacA_fam; 1.
DR   PROSITE; PS00450; ACONITASE_1; 1.
DR   PROSITE; PS01244; ACONITASE_2; FALSE_NEG.
KW   Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.
FT   METAL       295    295       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       353    353       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       356    356       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   415 AA;  44747 MW;  2B28CA0921842473 CRC64;

  Query Match            44.4%;  Score 44;  DB 1;  Length 415;
  Best Local Similarity  53.8%;  Pred. No. 28;
  Matches    7;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLG 13
              | |  |: |:|||
Db          2 PTWTEYILYKKLG 14
```

```
Search completed: October  2, 2003, 14:34:09
Job time : 23 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 14:30:36 ; Search time 94 Seconds
(without alignments)
43.924 Million cell updates/sec

Title: US-09-763-848-1
Perfect score: 99
Sequence: 1 PEWPSYLGYEKLGPYY 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	99	100.0	284	11	Q63759	Q63759 rattus norv
2	99	100.0	284	11	Q8C767	Q8c767 mus musculu
3	99	100.0	285	4	Q9H812	Q9h812 homo sapien
4	56	56.6	530	11	Q8BPV9	Q8bpv9 mus musculu
5	52	52.5	140	12	Q8JNS5	Q8jns5 human rhino
6	52	52.5	406	10	Q43509	Q43509 lycopersico
7	51	51.5	700	6	Q9GLG1	Q9glg1 macaca fasc
8	50	50.5	449	11	Q9ER53	Q9er53 mus musculu
9	50	50.5	462	11	Q9ER54	Q9er54 mus musculu
10	50	50.5	502	11	Q9ER55	Q9er55 mus musculu
11	50	50.5	720	11	Q9ER56	Q9er56 mus musculu
12	50	50.5	847	2	Q9ADT8	Q9adt8 salmonella
13	49	49.5	842	5	Q9U3A8	Q9u3a8 caenorhabdi
14	49	49.5	1160	5	P90935	P90935 caenorhabdi
15	49	49.5	1286	5	P90936	P90936 caenorhabdi
16	48	48.5	466	3	O00096	O00096 talaromyces
17	47.5	48.0	273	17	O57749	O57749 pyrococcus
18	47	47.5	186	2	Q9F7D5	Q9f7d5 salmonella
19	47	47.5	218	16	Q8EB78	Q8eb78 shewanella
20	47	47.5	245	16	Q8Z6K4	Q8z6k4 salmonella
21	47	47.5	245	16	Q9Z4S4	Q9z4s4 salmonella
22	47	47.5	301	5	Q9GYC3	Q9gyc3 leishmania
23	47	47.5	456	10	Q9LX95	Q9lx95 arabidopsis
24	47	47.5	486	10	Q9FKQ1	Q9fkq1 arabidopsis
25	47	47.5	566	5	O16200	O16200 caenorhabdi
26	47	47.5	703	11	Q64698	Q64698 rattus norv
27	46	46.5	76	2	Q8L236	Q8l236 salmonella
28	46	46.5	327	16	Q9I0M9	Q9i0m9 pseudomonas
29	46	46.5	381	11	Q91UZ9	Q91uz9 mus musculu
30	46	46.5	469	2	Q8GGL7	Q8ggl7 streptomyce
31	46	46.5	549	17	Q8TWI2	Q8twi2 methanopyru
32	46	46.5	703	11	Q91VA3	Q91va3 mus musculu
33	45.5	46.0	361	16	Q8F776	Q8f776 leptospira
34	45.5	46.0	739	2	O87381	O87381 haemophilus
35	45	45.5	344	2	O66257	O66257 actinobacil
36	45	45.5	385	10	O23955	O23955 gossypium h
37	45	45.5	528	2	Q8KPV4	Q8kpv4 synechococc
38	45	45.5	551	2	Q9RH54	Q9rh54 pantoea agg
39	45	45.5	724	13	Q8UW96	Q8uw96 xenopus lae
40	45	45.5	1227	16	Q9EWF3	Q9ewf3 streptomyce
41	44.5	44.9	143	2	Q54526	Q54526 streptomyce
42	44.5	44.9	145	2	Q54808	Q54808 streptomyce
43	44.5	44.9	161	2	Q55215	Q55215 streptomyce
44	44.5	44.9	162	16	Q98P38	Q98p38 rhizobium l
45	44.5	44.9	177	17	Q9YAM1	Q9yam1 aeropyrum p

ALIGNMENTS

RESULT 1
Q63759
ID Q63759 PRELIMINARY; PRT; 284 AA.
AC Q63759;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Serine-threonine specific protein phosphatase, glycogen-binding (GL)
DE subunit (EC 3.1.3.16).
GN PPP1R4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=96085228; PubMed=7498521;
RA Doherty M.J., Moorhead G., Morrice N., Cohen P., Cohen P.T.;
RT "Amino acid sequence and expression of the hepatic glycogen-binding
RT (GL)-subunit of protein phosphatase-1.";
RL FEBS Lett. 375:294-298(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RA Cohen P.T.W.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Y18208; CAA77083.1; -.
DR InterPro; IPR005036; CBM_21.
DR Pfam; PF03370; CBM_21; 1.
KW Hydrolase.
SQ SEQUENCE 284 AA; 32557 MW; 8FD4448CCF0EFE83 CRC64;

Query Match 100.0%; Score 99; DB 11; Length 284;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPYY 16
||||||||||||||||

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 14:32:16 ; Search time 29 Seconds
(without alignments)
23.344 Million cell updates/sec

Title: US-09-763-848-1
Perfect score: 99
Sequence: 1 PEWPSYLGYEKLGPYY 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	99	100.0	284	2	US-08-767-096-3	Sequence 3, Appli
2	99	100.0	284	3	US-09-480-203-3	Sequence 3, Appli
3	56	56.6	700	4	US-09-308-345A-46	Sequence 46, Appl
4	51	51.5	700	1	US-08-726-525-7	Sequence 7, Appli
5	51	51.5	700	2	US-08-487-942-7	Sequence 7, Appli
6	51	51.5	700	2	US-08-726-036A-7	Sequence 7, Appli
7	51	51.5	700	3	US-09-422-869-23	Sequence 23, Appl
8	51	51.5	700	4	US-09-083-516-7	Sequence 7, Appli
9	48	48.5	19	3	US-08-993-165-12	Sequence 12, Appl
10	48	48.5	19	4	US-09-540-448-12	Sequence 12, Appl
11	48	48.5	19	4	US-09-243-640-10	Sequence 10, Appl
12	48	48.5	19	4	US-08-929-847-12	Sequence 12, Appl
13	48	48.5	21	3	US-08-993-165-13	Sequence 13, Appl
14	48	48.5	21	4	US-09-540-448-13	Sequence 13, Appl
15	48	48.5	21	4	US-09-243-640-11	Sequence 11, Appl
16	48	48.5	21	4	US-08-929-847-13	Sequence 13, Appl
17	48	48.5	26	6	5196510-15	Patent No. 5196510
18	48	48.5	27	6	5196510-12	Patent No. 5196510
19	48	48.5	283	4	US-09-636-499-9	Sequence 9, Appli
20	48	48.5	283	4	US-09-636-499-24	Sequence 24, Appl
21	48	48.5	450	4	US-09-044-718-9	Sequence 9, Appli
22	48	48.5	466	3	US-08-868-435-31	Sequence 31, Appl
23	48	48.5	466	4	US-08-744-231-31	Sequence 31, Appl
24	48	48.5	466	4	US-09-273-871A-13	Sequence 13, Appl
25	47	47.5	703	2	US-08-835-099A-1	Sequence 1, Appli
26	47	47.5	703	3	US-09-157-349-1	Sequence 1, Appli
27	47	47.5	703	3	US-09-422-869-27	Sequence 27, Appl
28	47	47.5	703	4	US-09-308-345A-49	Sequence 49, Appl
29	47	47.5	712	2	US-08-835-099A-2	Sequence 2, Appli
30	47	47.5	712	3	US-09-157-349-2	Sequence 2, Appli
31	46	46.5	485	4	US-09-252-991A-31436	Sequence 31436, A
32	45	45.5	735	3	US-08-539-205A-2	Sequence 2, Appli
33	45	45.5	735	4	US-09-392-163A-2	Sequence 2, Appli
34	44	44.4	168	4	US-09-328-352-4319	Sequence 4319, Ap
35	43	43.4	111	4	US-09-899-896-7	Sequence 7, Appli
36	42	42.4	123	4	US-09-247-155-103	Sequence 103, App
37	42	42.4	196	4	US-08-679-493A-185	Sequence 185, App
38	42	42.4	438	1	US-08-480-604A-23	Sequence 23, Appl
39	42	42.4	438	2	US-08-405-496A-23	Sequence 23, Appl
40	42	42.4	438	3	US-08-915-136-23	Sequence 23, Appl
41	42	42.4	462	1	US-08-480-604A-26	Sequence 26, Appl
42	42	42.4	462	2	US-08-405-496A-26	Sequence 26, Appl
43	42	42.4	462	3	US-08-915-136-26	Sequence 26, Appl
44	42	42.4	686	4	US-09-653-839-8	Sequence 8, Appli
45	42	42.4	702	4	US-09-653-839-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-08-767-096-3
; Sequence 3, Application US/08767096
; Patent No. 5939284
;  GENERAL INFORMATION:
;    APPLICANT:  Cohen, Patricia T.W.
;    APPLICANT:  Cohen, Philip
;    APPLICANT:  Young, Peter R.
;    TITLE OF INVENTION:  A Protein Phosphatase 1 Binding Protein,
;    TITLE OF INVENTION:  R5
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-2799
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/767,096
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schreck, Patricia A.
;      REGISTRATION NUMBER:  33,777
;      REFERENCE/DOCKET NUMBER:  ATG50033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5031
;      TELEFAX:  610-270-5090
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  284 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-767-096-3

  Query Match           100.0%;  Score 99;  DB 2;  Length 284;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPYY 16
              ||||||||||||||||
Db        269 PEWPSYLGYEKLGPYY 284
```

```
RESULT 2
US-09-480-203-3
; Sequence 3, Application US/09480203
; Patent No. 6297359
; GENERAL INFORMATION:
;  APPLICANT: Patricia T.W. Cohen
;  APPLICANT: Phillip Cohen
;  APPLICANT: Peter R. Young
;  TITLE OF INVENTION: A PROTEIN PHOSPHATASE 1 BINDING PROTEIN
;  TITLE OF INVENTION: R5
;  FILE REFERENCE: ATG-50033-2
;  CURRENT APPLICATION NUMBER: US/09/480,203
;  CURRENT FILING DATE: 2000-01-10
;  EARLIER APPLICATION NUMBER: 09/300,327
;  EARLIER FILING DATE: 1999-04-27
;  EARLIER APPLICATION NUMBER: 08/767,096
;  EARLIER FILING DATE: 1996-12-05
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-480-203-3

  Query Match           100.0%;  Score 99;  DB 3;  Length 284;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPYY 16
              ||||||||||||||||
Db        269 PEWPSYLGYEKLGPYY 284


RESULT 3
US-09-308-345A-46
; Sequence 46, Application US/09308345A
; Patent No. 6569665
; GENERAL INFORMATION:
;  APPLICANT: BOEHM, Thomas;
;  APPLICANT:  DEAR, Neil T.
;  TITLE OF INVENTION: No. 6569665el calpains, their preparation and use
;  FILE REFERENCE: 0050/47576
;  CURRENT APPLICATION NUMBER: US/09/308,345A
;  CURRENT FILING DATE:  1999-05-19
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: WordPerfect v. 6.1
; SEQ ID NO 46
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: mouse
US-09-308-345A-46

  Query Match           56.6%;  Score 56;  DB 4;  Length 700;
  Best Local Similarity  66.7%;  Pred. No. 1.4;
  Matches   10;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || |||::|||| 
Db         52 PALPSSLGYKELGPY 66


RESULT 4
US-08-726-525-7
; Sequence 7, Application US/08726525
; Patent No. 5789181
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL INTERLEUKIN-1 RECEPTOR
;    TITLE OF INVENTION:  INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
;    TITLE OF INVENTION:  BINDING
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,525
;      FILING DATE:  07-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,942
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5258
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-726-525-7

  Query Match           51.5%;  Score 51;  DB 1;  Length 700;
  Best Local Similarity  60.0%;  Pred. No. 8;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 5
US-08-487-942-7
; Sequence 7, Application US/08487942
; Patent No. 5817476
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL INTERLEUKIN-1 RECEPTOR
;    TITLE OF INVENTION:  INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
;    TITLE OF INVENTION:  BINDING
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,942
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5258
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-487-942-7

  Query Match             51.5%;  Score 51;  DB 2;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 8;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 6
US-08-726-036A-7
; Sequence 7, Application US/08726036A
; Patent No. 5981482
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL INTERLEUKIN-1 RECEPTOR
;    TITLE OF INVENTION:  INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
;    TITLE OF INVENTION:  BINDING
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,036A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5258
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-726-036A-7

  Query Match             51.5%;  Score 51;  DB 2;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 8;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 7
US-09-422-869-23
; Sequence 23, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
;  APPLICANT: POLONSKY, KENNETH S.
;  APPLICANT: HORIKAWA, YUKIO
;  APPLICANT: ODA, NAOHISA
;  APPLICANT: COX, NANCY J.
;  APPLICANT: SREENAN, SEAMUS
;  APPLICANT: ZHOU, YUN-PING
;  APPLICANT: OTANI, KENICHI
;  APPLICANT: HANIS, CRAIG L.
;  APPLICANT: BELL, GRAEME I.
;  TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
;  FILE REFERENCE: ARCD:307
;  CURRENT APPLICATION NUMBER: US/09/422,869
;  CURRENT FILING DATE: 1999-10-21
;  EARLIER APPLICATION NUMBER: 60/134,175
;  EARLIER FILING DATE: 1999-05-13
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Human
US-09-422-869-23

  Query Match             51.5%;  Score 51;  DB 3;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 8;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 8
US-09-083-516-7
; Sequence 7, Application US/09083516
; Patent No. 6300086
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL INTERLEUKIN-1 RECEPTOR
;    TITLE OF INVENTION:  INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
;    TITLE OF INVENTION:  BINDING
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/083,516
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,942
```

```
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5258
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-09-083-516-7

  Query Match             51.5%;  Score 51;  DB 4;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 8;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 9
US-08-993-165-12
; Sequence 12, Application US/08993165A
; Patent No. 6123923
; GENERAL INFORMATION:
;  APPLICANT: Unger, Evan C
;  APPLICANT: Wu, Yunqiu
;  TITLE OF INVENTION: Optacoustic Contrast Agents And Methods For Their Use
;  FILE REFERENCE: UNGR1224
;  CURRENT APPLICATION NUMBER: US/08/993,165A
;  CURRENT FILING DATE: 1997-12-18
;  NUMBER OF SEQ ID NOS: 40
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: No. 6123923el Sequence
US-08-993-165-12

  Query Match             48.5%;  Score 48;  DB 3;  Length 19;
  Best Local Similarity   61.5%;  Pred. No. 0.51;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 PSYLGYEKLGPYY 16
              |||  |:  |||| 
Db          3 PSYYRYDGAGPYY 15


RESULT 10
US-09-540-448-12
; Sequence 12, Application US/09540448
; Patent No. 6403056
; GENERAL INFORMATION:
;  APPLICANT:  Unger, Evan C.
;  TITLE OF INVENTION:  Charged Lipids and Uses For The Same
;  FILE REFERENCE:  UNGR1592
;  CURRENT APPLICATION NUMBER: US/09/540,448
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  08/925,353
;  PRIOR FILING DATE:  1997-09-08
;  NUMBER OF SEQ ID NOS:  37
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: No. 6403056el Sequence
US-09-540-448-12

  Query Match             48.5%;  Score 48;  DB 4;  Length 19;
  Best Local Similarity   61.5%;  Pred. No. 0.51;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 PSYLGYEKLGPYY 16
              |||  |:  |||| 
Db          3 PSYYRYDGAGPYY 15


RESULT 11
US-09-243-640-10
; Sequence 10, Application US/09243640
; Patent No. 6521211
; GENERAL INFORMATION:
;  APPLICANT:  Unger, Evan C
;  APPLICANT:  Shen, Dekang
;  APPLICANT:  Wu, Guanli
;  TITLE OF INVENTION:  No. 6521211el Methods Of Imaging And Treatment With Targeted
;  TITLE OF INVENTION:  Compositions
;  FILE REFERENCE:  DUP-0463
;  CURRENT APPLICATION NUMBER: US/09/243,640
;  CURRENT FILING DATE:  1999-02-03
;  PRIOR APPLICATION NUMBER:  08/660,032
;  PRIOR FILING DATE:  1996-06-06
;  PRIOR APPLICATION NUMBER:  08/640,464
;  PRIOR FILING DATE:  1996-05-01
;  PRIOR APPLICATION NUMBER:  08/497,684
;  PRIOR FILING DATE:  1995-06-07
;  PRIOR APPLICATION NUMBER:  09/218,660
;  PRIOR FILING DATE:  1998-12-22
;  PRIOR APPLICATION NUMBER:  60/073,913
;  PRIOR FILING DATE:  1998-02-06
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: No. 6521211el Sequence
US-09-243-640-10

  Query Match             48.5%;  Score 48;  DB 4;  Length 19;
  Best Local Similarity   61.5%;  Pred. No. 0.51;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 PSYLGYEKLGPYY 16
              |||  |:  |||| 
Db          3 PSYYRYDGAGPYY 15


RESULT 12
US-08-929-847-12
; Sequence 12, Application US/08929847
; Patent No. 6548047
; GENERAL INFORMATION:
;  APPLICANT: Unger, Evan C.
;  TITLE OF INVENTION: Thermal Preactivation Of Gaseous Precursor Filled Compositions
;  FILE REFERENCE: BMS0441
;  CURRENT APPLICATION NUMBER: US/08/929,847
;  CURRENT FILING DATE:  1997-09-15
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
```

; FEATURE:
; OTHER INFORMATION: Completely synthetic sequence
US-08-929-847-12

Query Match 48.5%; Score 48; DB 4; Length 19;
Best Local Similarity 61.5%; Pred. No. 0.51;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PSYLGYEKLGPYY 16
||| |: ||||
Db 3 PSYYRYDGAGPYY 15

RESULT 13
US-08-993-165-13
; Sequence 13, Application US/08993165A
; Patent No. 6123923
; GENERAL INFORMATION:
; APPLICANT: Unger, Evan C
; APPLICANT: Wu, Yunqiu
; TITLE OF INVENTION: Optacoustic Contrast Agents And Methods For Their Use
; FILE REFERENCE: UNGR1224
; CURRENT APPLICATION NUMBER: US/08/993,165A
; CURRENT FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: No. 6123923el Sequence
US-08-993-165-13

Query Match 48.5%; Score 48; DB 3; Length 21;
Best Local Similarity 61.5%; Pred. No. 0.56;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PSYLGYEKLGPYY 16
||| |: ||||
Db 5 PSYYRYDGAGPYY 17

RESULT 14
US-09-540-448-13
; Sequence 13, Application US/09540448
; Patent No. 6403056
; GENERAL INFORMATION:
; APPLICANT: Unger, Evan C.
; TITLE OF INVENTION: Charged Lipids and Uses For The Same
; FILE REFERENCE: UNGR1592
; CURRENT APPLICATION NUMBER: US/09/540,448
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 08/925,353
; PRIOR FILING DATE: 1997-09-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: No. 6403056el Sequence
US-09-540-448-13

Query Match 48.5%; Score 48; DB 4; Length 21;
Best Local Similarity 61.5%; Pred. No. 0.56;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PSYLGYEKLGPYY 16
||| |: ||||
Db 5 PSYYRYDGAGPYY 17

RESULT 15
US-09-243-640-11
; Sequence 11, Application US/09243640
; Patent No. 6521211
; GENERAL INFORMATION:
; APPLICANT: Unger, Evan C
; APPLICANT: Shen, Dekang
; APPLICANT: Wu, Guanli
; TITLE OF INVENTION: No. 6521211el Methods Of Imaging And Treatment With Targeted
; TITLE OF INVENTION: Compositions
; FILE REFERENCE: DUP-0463
; CURRENT APPLICATION NUMBER: US/09/243,640
; CURRENT FILING DATE: 1999-02-03
; PRIOR APPLICATION NUMBER: 08/660,032
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: 08/640,464
; PRIOR FILING DATE: 1996-05-01
; PRIOR APPLICATION NUMBER: 08/497,684
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/218,660
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/073,913
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: No. 6521211el Sequence
US-09-243-640-11

Query Match 48.5%; Score 48; DB 4; Length 21;
Best Local Similarity 61.5%; Pred. No. 0.56;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PSYLGYEKLGPYY 16
||| |: ||||
Db 5 PSYYRYDGAGPYY 17

Search completed: October 2, 2003, 14:37:14
Job time : 30 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 14:36:01 ; Search time 66 Seconds
(without alignments)
38.355 Million cell updates/sec

Title: US-09-763-848-1
Perfect score: 99
Sequence: 1 PEWPSYLGYEKLGPYY 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	99	100.0	282	9	US-09-737-149-43	Sequence 43, Appl
2	99	100.0	284	9	US-09-737-149-42	Sequence 42, Appl
3	99	100.0	285	9	US-09-737-149-12	Sequence 12, Appl
4	53	53.5	545	15	US-10-156-761-9468	Sequence 9468, Ap
5	51	51.5	700	9	US-09-840-707A-9	Sequence 9, Appli
6	51	51.5	700	10	US-09-768-877-23	Sequence 23, Appl
7	51	51.5	700	11	US-09-884-319-7	Sequence 7, Appli
8	51	51.5	700	15	US-10-038-557A-9	Sequence 9, Appli
9	51	51.5	700	15	US-10-116-519-12	Sequence 12, Appl
10	48	48.5	19	15	US-10-046-801-12	Sequence 12, Appl
11	48	48.5	21	15	US-10-046-801-13	Sequence 13, Appl
12	48	48.5	283	15	US-10-229-358-9	Sequence 9, Appli
13	48	48.5	283	15	US-10-229-358-24	Sequence 24, Appl
14	48	48.5	450	15	US-10-062-848-9	Sequence 9, Appli
15	48	48.5	466	14	US-10-083-452-13	Sequence 13, Appl
16	47	47.5	647	15	US-10-116-519-10	Sequence 10, Appl
17	47	47.5	703	10	US-09-768-877-27	Sequence 27, Appl
18	46	46.5	469	12	US-10-314-657-71	Sequence 71, Appl
19	45	45.5	278	15	US-10-156-761-15067	Sequence 15067, A
20	45	45.5	385	10	US-09-837-751-4	Sequence 4, Appli
21	45	45.5	748	12	US-10-021-660-81	Sequence 81, Appl
22	44	44.4	255	15	US-10-224-446-30	Sequence 30, Appl
23	44	44.4	255	15	US-10-287-401-30	Sequence 30, Appl
24	43	43.4	111	9	US-09-899-896-7	Sequence 7, Appli
25	43	43.4	253	11	US-09-880-748-1835	Sequence 1835, Ap
26	43	43.4	300	9	US-09-794-960-5	Sequence 5, Appli
27	42	42.4	123	12	US-09-903-190-103	Sequence 103, App
28	42	42.4	343	9	US-09-794-960-4	Sequence 4, Appli
29	42	42.4	383	9	US-09-852-399-4	Sequence 4, Appli
30	42	42.4	399	15	US-10-224-446-4	Sequence 4, Appli
31	42	42.4	399	15	US-10-287-401-4	Sequence 4, Appli
32	42	42.4	414	12	US-10-097-111-290	Sequence 290, App
33	42	42.4	425	9	US-09-288-326-9	Sequence 9, Appli
34	42	42.4	434	11	US-09-910-186A-4	Sequence 4, Appli
35	42	42.4	435	11	US-09-910-186A-6	Sequence 6, Appli
36	42	42.4	437	11	US-09-910-186A-2	Sequence 2, Appli
37	42	42.4	702	15	US-10-116-519-11	Sequence 11, Appl
38	42	42.4	714	10	US-09-768-877-22	Sequence 22, Appl
39	42	42.4	714	15	US-10-116-519-6	Sequence 6, Appli
40	42	42.4	722	15	US-10-097-534-14	Sequence 14, Appl
41	42	42.4	1295	10	US-09-726-949A-1	Sequence 1, Appli
42	41.5	41.9	152	9	US-09-815-242-13962	Sequence 13962, A
43	41	41.4	1246	9	US-09-741-669-349	Sequence 349, App
44	41	41.4	1247	9	US-09-815-242-10145	Sequence 10145, A
45	41	41.4	1247	9	US-09-815-242-13841	Sequence 13841, A

ALIGNMENTS

RESULT 1
US-09-737-149-43
; Sequence 43, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus Sequence
; NAME/KEY: VARIANT
; LOCATION: (1)..(282)

```
;   OTHER INFORMATION: Where X is a residue at which the query and
;   OTHER INFORMATION: subject seuqnces are not identical.
US-09-737-149-43

  Query Match            100.0%;  Score 99;  DB 9;  Length 282;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPYY 16
              ||||||||||||||||
Db        267 PEWPSYLGYEKLGPYY 282


RESULT 2
US-09-737-149-42
; Sequence 42, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
;  APPLICANT: Spaderna, Steven K
;  APPLICANT:  Quinn, Kerry E.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Muralidhara, Padigaru
;  APPLICANT:  Spytek, Kimberly A.
;  TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
;  FILE REFERENCE: 15966-620 CIP
;  CURRENT APPLICATION NUMBER: US/09/737,149
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: 60/170,564
;  PRIOR FILING DATE: 1999-12-14
;  PRIOR APPLICATION NUMBER: 60/173,165
;  PRIOR FILING DATE: 1999-12-27
;  PRIOR APPLICATION NUMBER: 60/173,362
;  PRIOR FILING DATE: 1999-12-27
;  PRIOR APPLICATION NUMBER: 60/173,544
;  PRIOR FILING DATE: 1999-12-29
;  PRIOR APPLICATION NUMBER: 60/174,404
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: 60/174,962
;  PRIOR FILING DATE: 2000-01-07
;  PRIOR APPLICATION NUMBER: 60/223,929
;  PRIOR FILING DATE: 2000-08-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-09-737-149-42

  Query Match            100.0%;  Score 99;  DB 9;  Length 284;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPYY 16
              ||||||||||||||||
Db        269 PEWPSYLGYEKLGPYY 284


RESULT 3
US-09-737-149-12
; Sequence 12, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
;  APPLICANT: Spaderna, Steven K
;  APPLICANT:  Quinn, Kerry E.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Muralidhara, Padigaru
;  APPLICANT:  Spytek, Kimberly A.
;  TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
;  FILE REFERENCE: 15966-620 CIP
;  CURRENT APPLICATION NUMBER: US/09/737,149
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: 60/170,564
;  PRIOR FILING DATE: 1999-12-14
;  PRIOR APPLICATION NUMBER: 60/173,165
;  PRIOR FILING DATE: 1999-12-27
;  PRIOR APPLICATION NUMBER: 60/173,362
;  PRIOR FILING DATE: 1999-12-27
;  PRIOR APPLICATION NUMBER: 60/173,544
;  PRIOR FILING DATE: 1999-12-29
;  PRIOR APPLICATION NUMBER: 60/174,404
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: 60/174,962
;  PRIOR FILING DATE: 2000-01-07
;  PRIOR APPLICATION NUMBER: 60/223,929
;  PRIOR FILING DATE: 2000-08-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 285
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-737-149-12

  Query Match            100.0%;  Score 99;  DB 9;  Length 285;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPYY 16
              ||||||||||||||||
Db        270 PEWPSYLGYEKLGPYY 285


RESULT 4
US-10-156-761-9468
; Sequence 9468, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9468
;   LENGTH: 545
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-9468

  Query Match            53.5%;  Score 53;  DB 15;  Length 545;
  Best Local Similarity  62.5%;  Pred. No. 5.4;
  Matches   10;  Conservative    1;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 PEWPSYLGYEKLGPYY 16
              | ||  | ||:| |||
Db        128 PAWP--LSYEELEPYY 141


RESULT 5
US-09-840-707A-9
; Sequence 9, Application US/09840707A
; Patent No. US20020077276A1
; GENERAL INFORMATION:
;  APPLICANT: Fredeking, Terry M.
;  APPLICANT:  Ignatyev, George M.
```

```
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
;  TITLE OF INVENTION:  INFECTIONS AND OTHER DISORDERS
;  FILE REFERENCE: 24881-301C
;  CURRENT APPLICATION NUMBER: US/09/840,707A
;  CURRENT FILING DATE:  2001-04-23
;  PRIOR APPLICATION NUMBER: 09/562,979
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/198,210
;  PRIOR FILING DATE: 1999-04-27
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: IL-1 receptor intracellular ligand protein
;   OTHER INFORMATION: comprising amino acid sequence
;  PUBLICATION INFORMATION:
;   PATENT DOCUMENT NUMBER: 5,817,476
;   PATENT FILING DATE: 1995-06-07
;   PUBLICATION DATE: 1998-10-06
US-09-840-707A-9

  Query Match             51.5%;  Score 51;  DB 9;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 14;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 6
US-09-768-877-23
; Sequence 23, Application US/09768877
; Patent No. US20020150896A1
; GENERAL INFORMATION:
;  APPLICANT: POLONSKY, KENNETH S.
;  APPLICANT:  HORIKAWA, YUKIO
;  APPLICANT:  ODA, NAOHISA
;  APPLICANT:  COX, NANCY J.
;  APPLICANT:  SREENAN, SEAMUS
;  APPLICANT:  ZHOU, YUN-PING
;  APPLICANT:  OTANI, KENICHI
;  APPLICANT:  HANIS, CRAIG L.
;  APPLICANT:  BELL, GRAEME I.
;  TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
;  FILE REFERENCE: ARCD:307
;  CURRENT APPLICATION NUMBER: US/09/768,877
;  CURRENT FILING DATE:  2001-01-23
;  PRIOR APPLICATION NUMBER: 09/422,869
;  PRIOR FILING DATE: 1999-10-21
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Human
US-09-768-877-23

  Query Match             51.5%;  Score 51;  DB 10;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 14;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 7
US-09-884-319-7
; Sequence 7, Application US/09884319
; Publication No. US20030124625A1
;   GENERAL INFORMATION:
;        APPLICANT: Lin, Lih-Ling
;                   Graham, James
;        TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
;                            INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
;                            BINDING
;        NUMBER OF SEQUENCES: 7
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
;             STREET: 87 CambridgePark Drive
;             CITY: Cambridge
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02140
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/884,319
;             FILING DATE: 18-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/083,516
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/487,942
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Brown, Scott A.
;             REGISTRATION NUMBER: 32,724
;             REFERENCE/DOCKET NUMBER: GI5258
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 498-8224
;             TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 700 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-884-319-7

  Query Match             51.5%;  Score 51;  DB 11;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 14;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 8
US-10-038-557A-9
; Sequence 9, Application US/10038557A
; Publication No. US20030092684A1
; GENERAL INFORMATION:
;  APPLICANT: Fredeking, Terry M.
;  APPLICANT:  Ignatyev, George M.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
;  TITLE OF INVENTION:  INFECTIONS AND OTHER DISORDERS
;  FILE REFERENCE: 24881-301D
;  CURRENT APPLICATION NUMBER: US/10/038,557A
;  CURRENT FILING DATE:  2002-06-05
;  PRIOR APPLICATION NUMBER: 09/840,707
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: 09/562,979
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/198,210
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 14:23:05 ; Search time 82 Seconds
(without alignments)
30.971 Million cell updates/sec

Title: US-09-763-848-1
Perfect score: 99
Sequence: 1 PEWPSYLGYEKLGPYY 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	99	100.0	284	23	AAE14236	Human protein phos
2	99	100.0	285	22	AAU02201	Phosphatase 1 prot
3	99	100.0	285	22	AAB95633	Human protein sequ
4	90	90.9	16	21	AAY79064	C-terminal peptide
5	52	52.5	700	22	AAB86130	Rat calpain 80kDa
6	51	51.5	144	22	AAU23305	Novel human enzyme
7	51	51.5	700	18	AAW19992	Human CANP used to
8	51	51.5	700	21	AAB37797	Human interleukin-
9	51	51.5	700	22	AAB86128	Human calpain 80kD
10	51	51.5	700	23	AAE25059	Human calpain prot
11	50	50.5	447	23	AAM49720	Murine capn12 prot
12	50	50.5	462	23	AAM49719	Murine calpain pro
13	50	50.5	518	23	AAM49718	Murine calpain pro
14	50	50.5	720	23	AAM49717	Murine calpain pro
15	48.5	49.0	82	23	ABP10799	Human ORFX protein
16	48	48.5	19	18	AAW45497	Targeting ligand d
17	48	48.5	19	21	AAB20595	Glycoprotein GPIIb
18	48	48.5	19	23	ABG30388	Glycoprotein GpIIb
19	48	48.5	21	18	AAW45498	Targeting ligand d
20	48	48.5	21	21	AAB20596	Glycoprotein GPIIb
21	48	48.5	21	23	ABG30389	Glycoprotein GpIIb
22	48	48.5	26	12	AAR15276	Anti-thrombic fusi
23	48	48.5	27	12	AAR15273	Anti-thrombic fusi
24	48	48.5	127	11	AAR06355	Peptide correspond
25	48	48.5	219	15	AAR56235	h66-118/h13-65/l1-
26	48	48.5	223	15	AAR56236	l66-111/l12-65/h1-
27	48	48.5	283	22	AAB71730	P.piceum phytase p
28	48	48.5	283	22	AAB71733	P.piceum phytase p
29	48	48.5	443	21	AAB20512	Talaromyces thermo
30	48	48.5	443	21	AAY69555	Talaromyces thermo
31	48	48.5	466	19	AAW84355	Talaromyces thermo
32	48	48.5	466	20	AAY39899	T. thermophilus ph
33	47	47.5	300	22	AAM39363	Human polypeptide
34	47	47.5	392	21	AAG47099	Arabidopsis thalia
35	47	47.5	399	22	AAM41149	Human polypeptide
36	47	47.5	429	21	AAG47098	Arabidopsis thalia
37	47	47.5	486	21	AAG47097	Arabidopsis thalia
38	47	47.5	703	18	AAW41564	Human calpain. Ho
39	47	47.5	703	23	AAE14338	Human protease PRT
40	47	47.5	703	23	AAU72884	Human aspartyl pro
41	47	47.5	712	18	AAW41565	Human calpain. Ho
42	46	46.5	306	24	ABR41522	Human DITHP protei
43	46	46.5	467	23	AAE25058	Human calpain prot
44	46	46.5	469	24	ABU11410	Protein encoded by
45	46	46.5	648	23	ABG70268	Human Calpain-like

ALIGNMENTS

RESULT 1
AAE14236
ID AAE14236 standard; Protein; 284 AA.
XX
AC AAE14236;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human protein phosphatase-1 (PP-1).
XX
KW Human; protein phosphatase-1; PP-1; gene therapy; tranquilliser; amnesia;
KW arteriosclerosis; atherosclerosis; anxiety; anaemia; hepatitis; cataract;
KW amyotrophic lateral sclerosis; adenocarcinoma; cerebral palsy; psoriasis;
KW ulcerative colitis; myasthenia gravis; infection; schizophrenic disorder;
KW neurological disorder; epilepsy; neoplasm; Alzheimer's disease; dementia;
KW thyroiditis; dermatitis; diabetic mellitus; rheumatoid arthritis; stroke;
KW granulomatous disease; haemolytic anaemia; Crohn's disease; cancer; SCID;
KW severe combined immunodeficiency disease; immune system disorder; trauma;
KW developmental disorder; cell proliferative disorder; Addison's disease;
KW systemic lupus erythematosus; Parkinson's disease; myelofibrosis; AIDS;
KW leukaemia; antiinflammatory; cirrhosis; muscular dystrophy; allergy.
XX
OS Homo sapiens.
XX
PN WO200181590-A2.
XX
PD 01-NOV-2001.
XX
PF 19-APR-2001; 2001WO-US12902.
XX
PR 20-APR-2000; 2000US-199010P.
PR 05-MAY-2000; 2000US-202340P.

PR 10-MAY-2000; 2000US-203424P.
PR 18-MAY-2000; 2000US-205642P.
PR 02-JUN-2000; 2000US-208854P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Yue H, Khan FA, Wang YE, Patterson C, Gandhi AR;
PI Walia NK, Stewart EA, Tribouley CM, Hafalia A, Nguyen DB;
PI Elliott VS, Lee EA;
XX
DR WPI; 2002-034445/04.
DR N-PSDB; AAD23604.
XX
PT Novel polypeptides for diagnosing, preventing, treating immune system,
PT neurological, developmental and cell proliferative disorders including
PT cancer, comprises protein phosphatase polypeptides and encoding
PT polynucleotides -
XX
PS Claim 1; Page 97-98; 105pp; English.
XX
CC The invention relates to human protein phosphatases (PP-1 to PP-5) and
CC their corresponding DNA molecules. Protein phosphatases and their DNA's
CC are useful for diagnosis, treatment and prevention of immune system
CC disorders (AIDS, severe combined immunodeficiency disease (SCID), chronic
CC granulomatous disease, autoimmune haemolytic anaemia, Crohn's disease,
CC autoimmune thyroiditis, atopic dermatitis, diabetic mellitus, rheumatoid
CC arthritis, systemic lupus erythematosus, systemic sclerosis, Addison's
CC disease, ulcerative colitis, haemodialysis uveitis, myasthenia gravis,
CC trauma, viral, bacterial, fungal, parasitic, protozoal and helminthic
CC infections); neurological disorders (epilepsy, stroke, cerebral neoplasm,
CC Alzheimer's disease, Huntington's disease, Parkinson's disease,
CC amyotrophic lateral sclerosis, dementia, prion diseases); developmental
CC disorders (Down syndrome, cerebral palsy, spinal cord diseases, amnesia,
CC autonomic nervous system disorders, muscular dystrophy, anxiety, anaemia,
CC schizophrenic disorders, diabetic neuropathy, Tourette's disorder,
CC cataract, sensorineural hearing loss); cell proliferative disorders
CC (bursitis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis,
CC myelofibrosis, psoriasis, cancers including adenocarcinoma, leukaemia,
CC lymphoma, melanoma, myeloma, sarcoma and cancers of the bone, brain,
CC breast, ovary, bladder, heart, kidney, liver, and pancreas. Protein
CC phosphatases, its fragments and its antibodies are useful as elements on
CC a micro array which is useful to monitor or measure protein-protein
CC interactions, drug-target interactions and gene expression profiles.
CC The present sequence is human protein phosphatase-1 (PP-1).
XX
SQ Sequence 284 AA;

Query Match 100.0%; Score 99; DB 23; Length 284;
Best Local Similarity 100.0%; Pred. No. 6.5e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPYY 16
||||||||||||||||
Db 269 PEWPSYLGYEKLGPYY 284

RESULT 2
AAU02201
ID AAU02201 standard; Protein; 285 AA.
XX
AC AAU02201;
XX
DT 26-SEP-2001 (first entry)
XX
DE Phosphatase 1 protein-like protein, MEM6.
XX
KW Phosphatase 1; MEM1; therapeutic; diagnostic; MEM2;
KW human; Alzheimer's disease; Parkinson's Disease; cancer; nephrology;
KW female reproductive health; lung disorder; brain disorder; schizophrenia;
KW heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;
KW cobalamine deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;
KW vision-related disorder; neoplastic pathology; MEM7; MEM8.
XX
OS Homo sapiens.
XX
PN WO200144473-A2.
XX
PD 21-JUN-2001.
XX
PF 14-DEC-2000; 2000WO-US33909.
XX
PR 14-DEC-1999; 99US-0170564.
PR 27-DEC-1999; 99US-0173165.
PR 27-DEC-1999; 99US-0173362.
PR 29-DEC-1999; 99US-0173544.
PR 04-JAN-2000; 2000US-9966564.
PR 09-AUG-2000; 2000US-0223929.
PR 13-DEC-2000; 2000US-9966565.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Spaderna SK, Quinn KE, Shimkets RA, Muralidhara P, Spytek KA;
XX
DR WPI; 2001-398154/42.
DR N-PSDB; AAS06337.
XX
PT Novel polypeptide comprising members of protein families (e.g.,
PT seven-pass transmembrane receptor proteins) according to presence of
PT domains and sequence relatedness are useful for treating or preventing,
PT e.g., Alzheimer's and Parkinson's -
XX
PS Claim 1; Fig 27; 162pp; English.
XX
CC The sequence represents the amino acid sequence of phosphatase 1
CC protein-like protein, MEM6, selected from a group (MEM1-MEM8) comprising
CC members of protein families according to the presence of domains and
CC sequence relatedness, e.g., seven-pass transmembrane receptor protein
CC (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5),
CC phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).
CC The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are
CC all useful for treating or preventing a pathology associated with (I)
CC comprising administering (I), (II), or (III) to a subject (preferably a
CC human). In addition, (I), (II), and (III) may be used to manufacture a
CC medicament for treating a syndrome associated with a human disease that
CC is associated with (I). Furthermore, (I) may be used to identify agents
CC that bind to it, screen modulators of its activity and determine the
CC presence or predisposition to a disease associated with altered levels of
CC (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease,
CC cancer, nephrology, and female reproductive health. Disorders for MEM4
CC include those involving the lung and/or brain (e.g., schizophrenia). For
CC MEM5, disorders include heart (arrhythmic disorders) and other muscular
CC disorders, clotting deficiencies and cobalamine deficiencies (e.g.,
CC pernicious anaemia). Such disorders for MEM6 include diabetes, whereas
CC disorders for MEM7 and MEM8 include vision-related disorders, cancer,
CC and other neoplastic pathologies.
XX
SQ Sequence 285 AA;

Query Match 100.0%; Score 99; DB 22; Length 285;
Best Local Similarity 100.0%; Pred. No. 6.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPYY 16
||||||||||||||||
Db 270 PEWPSYLGYEKLGPYY 285

RESULT 3
AAB95633
ID AAB95633 standard; Protein; 285 AA.
XX
AC AAB95633;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18363.

XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 18363; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 285 AA;

Query Match 100.0%; Score 99; DB 22; Length 285;
Best Local Similarity 100.0%; Pred. No. 6.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPYY 16
||||||||||||||||
Db 270 PEWPSYLGYEKLGPYY 285

RESULT 4
AAY79064
ID AAY79064 standard; peptide; 16 AA.
XX
AC AAY79064;
XX
DT 12-JUN-2000 (first entry)
XX
DE C-terminal peptide of rat liver PP1 GL subunit.
XX
KW Protein phosphatase 1; PP1; glycogen targetting subunit; blood glucose;
KW phosphorylase alpha; glycogen synthesis; hyperglycaemic disorder; GL;
KW type I diabetes; type II diabetes.
XX
OS Rattus sp.
XX
PN WO200012549-A1.
XX
PD 09-MAR-2000.
XX
PF 19-AUG-1999; 99WO-GB02761.
XX
PR 27-AUG-1998; 98GB-0018650.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Cohen PTW, Armstrong CG, Doherty MJ;
XX
DR WPI; 2000-256587/22.
XX
PT Lowering blood sugar levels in the treatment of diabetes, using a
PT compound that blocks interaction between phosphorylase alpha and
PT protein phosphatase 1 glycogen-targetting subunit -
XX
PS Claim 4; Page 35; 51pp; English.
XX
CC This sequence represents the C-terminal amino acid sequence of the rat
CC liver protein phosphatase 1 (PP1) glycogen targetting subunit (GL). The
CC invention relates to the medicinal use of a compound capable of blocking
CC the interaction of phosphorylase alpha with GL, where the compound
CC comprises the present peptide sequence or a fragment of it. When
CC phosphorylase alpha binds to GL it potently inhibits its glycogen
CC synthase phosphatase activity and inhibits glycogen synthesis, this
CC contributes to high blood glucose levels. The invention also relates to
CC a method for identifying the compound. The compound can be used to reduce
CC the blood glucose level of a mammal, particularly a human, in
CC hyperglycaemic disorders such as type I or type II diabetes.
XX
SQ Sequence 16 AA;

Query Match 90.9%; Score 90; DB 21; Length 16;
Best Local Similarity 93.8%; Pred. No. 6.8e-07;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPYY 16
||||||||||||| |||
Db 1 PEWPSYLGYEKLYPYY 16

RESULT 5
AAB86130
ID AAB86130 standard; protein; 700 AA.
XX
AC AAB86130;
XX
DT 27-JUL-2001 (first entry)
XX
DE Rat calpain 80kDa subdomain protein fragment.
XX
KW Calpain; calcium-activated cystein proteinase; human; spatial structure;
KW Ca-activated cystein proteinase; protein coordinate data; treatment;
KW structure-function study; ischemic condition; muscular dystrophy; tumor;
KW muscular; antitumor.
XX
OS Rattus norvegicus.
XX
PN EP1108779-A2.
XX
PD 20-JUN-2001.
XX
PF 13-DEC-2000; 2000EP-0127369.

XX
PR 14-DEC-1999; 99DE-1060225.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PROT-) PROTEROS BIOSTRUCTURES GMBH.
XX
PI Strobl S, Fernandez-Catalan C, Bode W, Huber R, Suzuki K;
XX
DR WPI; 2001-376928/40.
XX
PT Spatial structures containing calpain-derived polypeptides, useful for
PT identifying calpain modulators and substrates, potentially useful e.g.
PT as antitumor agents -
XX
PS Claim 15; Fig 6; 182pp; German.
XX
CC This invention describes the novel spatial structure of human and rat
CC neutral calcium-activated cystein protease (calpain) family. The spatial
CC structure (especially crystalline forms) are used for structure-function
CC studies, particularly for identifying (pseudo)substrates, inhibitors and
CC activators of calpains, potentially useful for treatment of ischemic
CC conditions, muscular dystrophy and/or tumors. The products of the
CC invention have anti-ischemic, muscular and antitumor activity. This
CC sequence represents the rat calpain 80kDa subunit described in the
CC method of the invention.
XX
SQ Sequence 700 AA;

Query Match 52.5%; Score 52; DB 22; Length 700;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PALPSSLGFKELGPY 66
```

RESULT 6
AAU23305
ID AAU23305 standard; Protein; 144 AA.
XX
AC AAU23305;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #391.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465566/50.
DR N-PSDB; AAS41175.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 11; SEQ ID No 1301; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 144 AA;

Query Match 51.5%; Score 51; DB 22; Length 144;
Best Local Similarity 60.0%; Pred. No. 5.5;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PEWPSYLGYEKLGPY 15
| || ||:::|||
Db 81 PAIPSALGFKELGPY 95

RESULT 7
AAW19992
ID AAW19992 standard; Protein; 700 AA.
XX
AC AAW19992;
XX
DT 27-AUG-1997 (first entry)
XX
DE Human CANP used to identify inhibitors of interleukin-1 activity.
XX
KW IL; interleukin; receptor; ligand; screening assay; inhibitor;
KW IL-1 mediated response; inflammation; inflammatory; antibody;
KW intracellular domain; CANP; calcium activated neutral protease.
XX
OS Homo sapiens.
XX
PN WO9640907-A1.
XX
PD 19-DEC-1996.
XX
PF 06-MAY-1996; 96WO-US06363.
XX
PR 07-JUN-1995; 95US-0487942.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Graham J, Lin L;
XX
DR WPI; 1997-052315/05.
XX
PT Interleukin-1 receptor intracellular ligand proteins and related DNA
PT - used to identify inhibitors of the proteins for treatment of
PT inflammation
XX
PS Claim 14; Page 36-38; 54pp; English.
XX
CC AAW19992 represents human calcium activated neutral protease (CANP).
CC This protein was found to have an area of high homology with an
CC interleukin-1 receptor (IL-1-R) intracellular ligand (encoded by cDNA
CC clone 14w, see AAT71218) and thus will display some of the same
CC properties of this protein. IL-1-R intracellular ligand proteins are
CC used to screen for agents (e.g. antibodies) that are capable of
CC inhibiting or blocking the binding of an IL-1-R intracellular ligand
CC to the intracellular domain of IL-1-R, i.e. inhibitors of IL-1
CC activity. Such agents can be used to treat inflammatory conditions.
XX
SQ Sequence 700 AA;

```
  Query Match             51.5%;  Score 51;  DB 18;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 30;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 8
AAB37797
ID   AAB37797 standard; Protein; 700 AA.
XX
AC   AAB37797;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Human interleukin-1 receptor intracellular ligand protein #4.
XX
KW   Human; interleukin-1; IL-1; IL-1alpha; IL-1beta; IL-1 receptor;
KW   antiinflammatory; haemostatic; antibacterial; immunosuppressive;
KW   immunomodulator; cardiant; cytostatic; neuroprotective; respiratory;
KW   inflammation; infection; sepsis; cachexia; autoimmune disorder;
KW   cardiovascular disorder; chronic myelogenous leukaemia;
KW   multiple sclerosis; inflammatory bowel disease; Crohn's disease.
XX
OS   Homo sapiens.
XX
PN   WO200064479-A1.
XX
PD   02-NOV-2000.
XX
PF   26-APR-2000; 2000WO-US11700.
XX
PR   27-APR-1999;   99US-0301274.
XX
PA   (ANTI-) ANTIBODY SYSTEMS INC.
XX
PI   Fredeking TM,  Ignatyev GM;
XX
DR   WPI; 2000-679646/66.
XX
PT   Novel compositions comprising tetracycline or tetracycline-like
PT   compounds for the treatment and/or prevention of acute inflammatory
PT   responses and diseases, e.g. septic shock and immune complex-induced
PT   colitis -
XX
PS   Disclosure; Page 159-162; 183pp; English.
XX
CC   The present sequence is given in a specification relating to novel
CC   compositions and methods containing tetracycline or tetracycline-like
CC   compounds for treating and/or preventing acute inflammatory responses and
CC   diseases. Such diseases include acute inflammatory conditions associated
CC   with viral haemorrhagic diseases (including diseases caused by
CC   Bunyaviridea, Filoviridae, Flaviviridae or Arenaviridae viruses),
CC   parasitic diseases, bacterial infections, sepsis, cachexia, autoimmune
CC   disorders, acute cardiovascular events, chronic myelogenous leukaemia and
CC   transplanted bone marrow-induced graft-versus-host disease, septic shock,
CC   immune complex-induced colitis, cerebrospinal fluid inflammation,
CC   multiple sclerosis, inflammatory responses associated with trauma,
CC   systemic inflammatory response syndrome (SIRS), adult respiratory
CC   distress syndrome (ARDS), acute liver failure, inflammatory bowel disease
CC   and Crohn's disease.
XX
SQ   Sequence   700 AA;

  Query Match             51.5%;  Score 51;  DB 21;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 30;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 9
AAB86128
ID   AAB86128 standard; protein; 700 AA.
XX
AC   AAB86128;
XX
DT   27-JUL-2001  (first entry)
XX
DE   Human calpain 80kDa subdomain protein fragment.
XX
KW   Calpain; calcium-activated cystein proteinase; human; spatial structure;
KW   Ca-activated cystein proteinase; protein coordinate data; treatment;
KW   structure-function study; ischemic condition; muscular dystrophy; tumor;
KW   muscular; antitumor.
XX
OS   Homo sapiens.
XX
PN   EP1108779-A2.
XX
PD   20-JUN-2001.
XX
PF   13-DEC-2000; 2000EP-0127369.
XX
PR   14-DEC-1999;   99DE-1060225.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (PROT-) PROTEROS BIOSTRUCTURES GMBH.
XX
PI   Strobl S,  Fernandez-Catalan C,  Bode W,  Huber R,  Suzuki K;
XX
DR   WPI; 2001-376928/40.
XX
PT   Spatial structures containing calpain-derived polypeptides, useful for
PT   identifying calpain modulators and substrates, potentially useful e.g.
PT   as antitumor agents -
XX
PS   Claim 15; Fig 4; 182pp; German.
XX
CC   This invention describes the novel spatial structure of human and rat
CC   neutral calcium-activated cystein protease (calpain) family. The spatial
CC   structure (especially crystalline forms) are used for structure-function
CC   studies, particularly for identifying (pseudo)substrates, inhibitors and
CC   activators of calpains, potentially useful for treatment of ischemic
CC   conditions, muscular dystrophy and/or tumors. The products of the
CC   invention have anti-ischemic, muscular and antitumor activity. This
CC   sequence represents the human calpain 80kDa subunit described in the
CC   method of the invention.
XX
SQ   Sequence   700 AA;

  Query Match             51.5%;  Score 51;  DB 22;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 30;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGPY 15
              |  || ||:::|||| 
Db         52 PAIPSALGFKELGPY 66


RESULT 10
AAE25059
ID   AAE25059 standard; Protein; 700 AA.
XX
AC   AAE25059;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human calpain protein #2.
XX
KW   Human; calpain; nervous system disorder; amyotrophic lateral sclerosis;
KW   Parkinson's disease; dementia; genito-urinary system disorder; stroke;
```

KW Alzheimer's disease; multiple sclerosis; benign prostate hyperplasia;
KW urinary incontinence; gene therapy; cytostatic; nootropic; uropathic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
PN WO200248326-A2.
XX
PD 20-JUN-2002.
XX
PF 14-DEC-2001; 2001WO-EP14819.
XX
PR 14-DEC-2000; 2000US-255058P.
XX
PA (FARB) BAYER AG.
XX
PI Ramakrishnan S;
XX
DR WPI; 2002-537625/57.
XX
PT New human calpain polypeptide, useful for treating peripheral and
PT central nervous system disorder and genito-urinary system disorders
PT including urinary incontinence and benign prostate hyperplasia -
XX
PS Disclosure; Page 101-104; 110pp; English.
XX
CC The invention relates to novel human calpain proteins and polynucleotides
CC encoding such proteins. Calpain sequences of the invention are useful for
CC treating, ameliorating or correcting dysfunctions or diseases such as
CC peripheral or central nervous system (CNS) disorders (e.g., Parkinson's
CC disease, Alzheimer's disease, multiple sclerosis, stroke, amyotrophic
CC lateral sclerosis, dementia) and genito-urinary system disorders such
CC as urinary incontinence and benign prostate hyperplasia. They are also
CC used in gene therapy. The present sequence is human calpain protein.
XX
SQ Sequence 700 AA;

Query Match 51.5%; Score 51; DB 23; Length 700;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```
Qy      -       1 PEWPSYLGYEKLGPY 15
                  |  || ||:::|||| 
Db             52 PAIPSALGFKELGPY 66
```

RESULT 11
AAM49720
ID AAM49720 standard; Protein; 447 AA.
XX
AC AAM49720;
XX
DT 11-JUN-2002 (first entry)
XX
DE Murine capn12 protein.
XX
KW Calpain protease; capn12; splice variant; murine; gene therapy;
KW screening; diagnosis.
XX
OS Mus sp.
XX
PN DE10031932-A1.
XX
PD 10-JAN-2002.
XX
PF 30-JUN-2000; 2000DE-1031932.
XX
PR 30-JUN-2000; 2000DE-1031932.
XX
PA (BADI) BASF AG.
XX
PI Not given;
XX
DR WPI; 2002-115441/16.
DR N-PSDB; ABA99771.
XX
PT New calpain protein 12 with cysteine protease activity, useful for
PT treating specific deficiency disorders -
XX
PS Claim 2; Page 18-20; 36pp; German.
XX
CC This invention describes a novel murine calpain protease 12 (capn12).
CC The calpain protease of the invention, related proteins and nucleic acid
CC that encodes it, are useful for treatment (including gene therapy) of
CC diseases associated with insufficient expression of the calpain protease.
CC The protein is also used to screen for calpain protein effectors and to
CC raise specific immunoglobulins (Ig) useful for diagnosis. Also the
CC polynucleotide encoding capn12 is useful, e.g. as primers and probes, for
CC diagnosis of diseases, or predisposition to them, and for recombinant
CC production of capn12. This sequence represents the calpain protease
CC capn12 protein described in the disclosure of the invention.
XX
SQ Sequence 447 AA;

Query Match 50.5%; Score 50; DB 23; Length 447;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 PEWPSYLGYEKLGP 14
              |  |  |||:||||
Db         52 PAGPDALGYDKLGP 65
```

RESULT 12
AAM49719
ID AAM49719 standard; Protein; 462 AA.
XX
AC AAM49719;
XX
DT 11-JUN-2002 (first entry)
XX
DE Murine calpain protease 12 variant capn12C.
XX
KW Calpain protease; capn12; splice variant; murine; gene therapy;
KW screening; diagnosis; capn12C.
XX
OS Mus sp.
XX
PN DE10031932-A1.
XX
PD 10-JAN-2002.
XX
PF 30-JUN-2000; 2000DE-1031932.
XX
PR 30-JUN-2000; 2000DE-1031932.
XX
PA (BADI) BASF AG.
XX
PI Not given;
XX
DR WPI; 2002-115441/16.
DR N-PSDB; ABA99770.
XX
PT New calpain protein 12 with cysteine protease activity, useful for
PT treating specific deficiency disorders -
XX
PS Claim 2; Page 17-18; 36pp; German.
XX
CC This invention describes a novel murine calpain protease 12 (capn12).
CC The calpain protease of the invention, related proteins and nucleic acid
CC that encodes it, are useful for treatment (including gene therapy) of
CC diseases associated with insufficient expression of the calpain protease.
CC The protein is also used to screen for calpain protein effectors and to
CC raise specific immunoglobulins (Ig) useful for diagnosis. Also the
CC polynucleotide encoding capn12 is useful, e.g. as primers and probes, for
CC diagnosis of diseases, or predisposition to them, and for recombinant

```
CC   production of capn12. This sequence represents the calpain protease
CC   splice variant capn12C described in the disclosure of the invention.
XX
SQ   Sequence   462 AA;

  Query Match             50.5%;  Score 50;  DB 23;  Length 462;
  Best Local Similarity   64.3%;  Pred. No. 27;
  Matches    9;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGP 14
              |  |  |||:||||
Db         52 PAGPDALGYDKLGP 65


RESULT 13
AAM49718
ID   AAM49718 standard; Protein; 518 AA.
XX
AC   AAM49718;
XX
DT   11-JUN-2002  (first entry)
XX
DE   Murine calpain protease 12 variant capn12B.
XX
KW   Calpain protease; capn12; splice variant; murine; gene therapy;
KW   screening; diagnosis; capn12B.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 493..510
FT                   /note= "Encoded by GGTGGT"
XX
PN   DE10031932-A1.
XX
PD   10-JAN-2002.
XX
PF   30-JUN-2000; 2000DE-1031932.
XX
PR   30-JUN-2000; 2000DE-1031932.
XX
PA   (BADI ) BASF AG.
XX
PI   Not given;
XX
DR   WPI; 2002-115441/16.
DR   N-PSDB; ABA99769.
XX
PT   New calpain protein 12 with cysteine protease activity, useful for
PT   treating specific deficiency disorders -
XX
PS   Claim 2; Page 15-17; 36pp; German.
XX
CC   This invention describes a novel murine calpain protease 12 (capn12).
CC   The calpain protease of the invention, related proteins and nucleic acid
CC   that encodes it, are useful for treatment (including gene therapy) of
CC   diseases associated with insufficient expression of the calpain protease.
CC   The protein is also used to screen for calpain protein effectors and to
CC   raise specific immunoglobulins (Ig) useful for diagnosis. Also the
CC   polynucleotide encoding capn12 is useful, e.g. as primers and probes, for
CC   diagnosis of diseases, or predisposition to them, and for recombinant
CC   production of capn12. This sequence represents the calpain protease
CC   splice variant capn12B described in the disclosure of the invention.
XX
SQ   Sequence   518 AA;

  Query Match             50.5%;  Score 50;  DB 23;  Length 518;
  Best Local Similarity   64.3%;  Pred. No. 31;
  Matches    9;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGP 14
              |  |  |||:||||
Db         52 PAGPDALGYDKLGP 65


RESULT 14
AAM49717
ID   AAM49717 standard; Protein; 720 AA.
XX
AC   AAM49717;
XX
DT   11-JUN-2002  (first entry)
XX
DE   Murine calpain protease 12 variant capn12A.
XX
KW   Calpain protease; capn12; splice variant; murine; gene therapy;
KW   screening; diagnosis; capn12A.
XX
OS   Mus sp.
XX
PN   DE10031932-A1.
XX
PD   10-JAN-2002.
XX
PF   30-JUN-2000; 2000DE-1031932.
XX
PR   30-JUN-2000; 2000DE-1031932.
XX
PA   (BADI ) BASF AG.
XX
PI   Not given;
XX
DR   WPI; 2002-115441/16.
DR   N-PSDB; ABA99768.
XX
PT   New calpain protein 12 with cysteine protease activity, useful for
PT   treating specific deficiency disorders -
XX
PS   Claim 2; Page 13-15; 36pp; German.
XX
CC   This invention describes a novel murine calpain protease 12 (capn12).
CC   The calpain protease of the invention, related proteins and nucleic acid
CC   that encodes it, are useful for treatment (including gene therapy) of
CC   diseases associated with insufficient expression of the calpain protease.
CC   The protein is also used to screen for calpain protein effectors and to
CC   raise specific immunoglobulins (Ig) useful for diagnosis. Also the
CC   polynucleotide encoding capn12 is useful, e.g. as primers and probes, for
CC   diagnosis of diseases, or predisposition to them, and for recombinant
CC   production of capn12. This sequence represents the calpain protease
CC   splice variant capn12A described in the disclosure of the invention.
XX
SQ   Sequence   720 AA;

  Query Match             50.5%;  Score 50;  DB 23;  Length 720;
  Best Local Similarity   64.3%;  Pred. No. 44;
  Matches    9;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 PEWPSYLGYEKLGP 14
              |  |  |||:||||
Db         52 PAGPDALGYDKLGP 65


RESULT 15
ABP10799
ID   ABP10799 standard; Protein; 82 AA.
XX
AC   ABP10799;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:21580.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
```

KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN26551.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 21580; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 82 AA;

Query Match 49.0%; Score 48.5; DB 23; Length 82;
Best Local Similarity 60.0%; Pred. No. 7.1;
Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

```
Qy          1 PEWPSYLG-YEKLGP 14
              | |||:|| ::| ||
Db         32 PAWPSHLGNHKKRGP 46
```

Search completed: October 2, 2003, 14:33:40
Job time : 83 secs